UNCLASSIFIED

AD NUMBER
AD357877

CLASSIFICATION CHANGES

TO:	UNCLASSIFIED
FROM:	SECRET

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited.

FROM:

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; DEC 1961. Other requests shall be referred to Adjutant General's Office (Army), Washington, DC 20310.

AUTHORITY

AGO D/A ltr 9 Nov 1973 ; AGO D/A ltr 9 Nov 1973

THIS PAGE IS UNCLASSIFIED

UNCLASSIFIED

AD 357877

UNCLASSIFIED

9. "Radar Reflectivity Measurements of the B-57 'Canberra' Aircraft at 3000 Mc (U)," Radiation, Inc., Report No. 1049-34, 30 July 1958. (AD 162 305). ~~SECRET~~ Unclassified

10. "A Study of Electronic Countermeasures for Application Against Air Defense Guided Missile Systems," The University of Michigan, Willow Run Research Center, Report No. UMM-11, 15 April 1953. SECRET

11. D. M. Grimes, J. A. Boyd, and M. H. Miller,
& "Basic Principles of Doppler Fuze Counter-
12. measures, Part IV," Tech. Report No. 4, Electronic Defense Group, Department of Electrical Engineering, The University of Michigan, November 1952. SECRET

13. T. S. Kuhn, "Theory of Ship Echoes as Applied to Naval RCM Operations," Report No. 411-93, Project RP-186, Harvard University, Radio Research Laboratory, 14 July 1944.

14. F. R. Bacon Jr., "VT Fuze Countermeasures Systems Analysis," Technical Report No. 65, The University of Michigan, Department of Electrical Engineering, Electronic Defense Group, April 1957. SECRET

AD-357 877

Page 13-22

12-61 kg

Electronic Countermeasures

SECRET

ELECTRONIC
COUNTERMEASURES [U]

Editors

ed. by **J. A. BOYD**,

D. B. HARRIS,

D. D. KING

and **H. W. WELCH, Jr.**

SECRET

EDITORS

Joseph A. Boyd
   *Director, Institute of Science and
   Technology
   The University of Michigan
   Ann Arbor, Michigan*

Donald B. Harris
   *Senior Executive Engineer
   Stanford Research Institute
   Menlo Park, California
           and
   Research Associate
   Stanford Electronics Laboratory
   Stanford University
   Stanford, California*

Donald D. King
   *Vice President, Research
   Electronic Communications, Inc.
   Research Division
   Timonium, Maryland*

H. W. Welch, Jr.
   *Director, Research and Development
   Military Electronics Division
   Motorola, Inc.
   Scottsdale, Arizona*

AUTHORS

A. L. Aden
Research and Development
Military Electronics Division
Motorola, Inc.
Scottsdale, Arizona

William E. Ayer
Applied Technology, Inc.
Palo Alto, California

Ben F. Barton
Department of Electrical Engineering
The University of Michigan
Ann Arbor, Michigan

K. H. Barney
Sperry Gyroscope Company
Great Neck, New York

R. H. Benninghof
Aircraft Armaments, Inc.
Cockeysville, Maryland

Theodore G. Birdsall
Department of Electrical Engineering
The University of Michigan
Ann Arbor, Michigan

Herman Blasbalg
Electronic Communications, Inc.
Timonium, Maryland

J. A. Boyd
Institute of Science and
    Technology
The University of Michigan
Ann Arbor, Michigan

M. E. Brodwin
Department of Electrical Engineering
The Technological Institute,
    Northwestern University
Evanston, Illinois

Paul W. Crapuchettes
Vacuum Tube Division, Litton
    Industries, Inc.
San Carlos, California

C. Burton Crumly
Microwave Tube Division
Zenith Radio Research Corporation
Menlo Park, California

Edward C. Dench
Spencer Laboratory,
Raytheon Manufacturing Company
Burlington, Massachusetts

L. A. deRosa
ITT Communication Systems, Inc.
Paramus, New Jersey

Marshall D. Earle
Aerospace Corporation
Los Angeles, California

W. A. Edson
Microwave Laboratory, General Electric
    Company
Palo Alto, California

L. W. Evans
Electronic Defense Laboratory, Sylvania
    Electric Products, Inc.
Mountain View, California

Hansford W. Farris
Department of Electrical Engineering
The University of Michigan
Ann Arbor, Michigan

Eugene G. Fubini
Office of Defense Research and
    Engineering
Washington, D. C.

A. T. Goble
Union College
Schenectady, New York

John V. N. Granger
Granger Associates
Palo Alto, California

A. E. Hałteman
Electronic Defense Laboratory, Sylvania
    Electronic Products, Inc.
Mountain View, California

Donald B. Harris
Stanford Research Institute
Menlo Park, California

vi

Joseph F. Hull
  Litton Industries
  San Carlos, California

Donald D. King
  Electronic Communications, Inc.
  Research Division
  Timonium, Maryland

G. L. Lansman
  Military Electronics Division
  Motorola, Inc.
  Scottsdale, Arizona

Lloyd K. Lauderdale
  The Johns Hopkins University Radiation
    Laboratory
  Baltimore, Maryland

H. O. Lorensen
  U. S. Naval Research Laboratory
  Washington, D. C.

Robert T. McCann
  Defense Electronics Division
  General Electric Company
  Syracuse, New York

Alan B. Macnee
  Department of Electrical Engineering
  The University of Michigan
  Ann Arbor, Michigan

Yuji Morita
  Institute of Science and Technology
  The University of Michigan
  Ann Arbor, Michigan

Richard B. Nelson
  Varian Associates
  Palo Alto, California

John M. Osepchuk
  Crossed-Field Tube Laboratories
  Raytheon Manufacturing Company
  Burlington, Massachusetts

J. M. Pettit
  College of Engineering
  Stanford University
  Stanford, California

William R. Rambo
  Stanford Electronics Laboratories
  Stanford University
  Stanford, California

Robert H. Richard
  The Johns Hopkins University Radiation
    Laboratory
  Baltimore, Maryland

Russell A. Rollin, Jr.
  Institute of Science and Technology
  The University of Michigan
  Ann Arbor, Michigan

Samuel Stiber
  U. S. Army Signal Research and
    Development Laboratory
  Fort Monmouth, New Jersey

Wilson P. Tanner, Jr.
  Department of Electrical Engineering
  The University of Michigan
  Ann Arbor, Michigan

R. K. Thomas
  Research Division
  Electronic Communications, Inc.
  Timonium, Maryland

Joseph H. Vogelman
  Dynamic Electronics-New York, Inc.
  New York, New York

Alan T. Waterman, Jr.
  Stanford Electronics
    Laboratory
  Stanford University
  Stanford, California

D. A. Watkins
  Department of Electrical Engineering
  Stanford University
  Stanford, California

H. W. Welch, Jr.
  Military Electronics Division
  Motorola, Inc.
  Scottsdale, Arizona

D. A. Wilbur
  Research Laboratory, General Electric
    Company
  Schenectady, New York

T. A. Wild
  Research Division
  Electronic Communications, Inc.
  Timonium, Maryland

vii

FOREWORD

The history of science shows, almost without exception, that new knowledge of any nature, no matter how acquired or with what objective, eventually finds useful applications. Such knowledge may be gained for its own sake, as in basic research programs. The results of basic research may become useful in many areas, the exact nature of which can seldom be ascertained at the time when the research is started. Similarly, applied research undertaken with more or less specific objectives in mind may find applications quite unrelated to the original objectives.

This is the case with the electronic countermeasures program. Since its inception on a large scale under the auspices of the National Defense Research Committee and the Armed Services during World War II, this program has resulted in the development of techniques applicable not only to the particular objectives originally established, but also to many other purposes not originally envisioned. Some basic research, for which generally useful applications were to be expected, has been done as a part of this program. In other cases, applied research projects originally directed toward specific ends yielded not only the results intended but also other techniques and knowledge of general usefulness.

It is partly on account of this aspect of the ECM program that this book has been written. It is primarily a textbook of electronic countermeasures techniques intended for use in teaching those engaged in work on the ECM program. In addition, it documents research and development results which can be used not only for countermeasures but for other purposes as well. For example, all the following material is of general interest: Chapter 6, "Intercept Probability and Receiver Parameters"; Chapter 7, "Detection and Analysis of Signals"; Chapter 9, "The Intercept Receiver"; Chapter 10, "Direction Finding"; and all of Part IV, "Components", including circuit techniques, various types of microwave tubes, ferroelectric and ferromagnetic devices, and propagation. On the other hand, Part III, which covers specific countermeasures equipment and techniques, will probably be most useful to those actively engaged in the ECM program.

Many of the authors started their association with electronic countermeasures during World War II at the Radio Research Laboratory of Harvard University, at the Airborne Instruments Laboratory, or at private

ix

or government laboratories engaged in ECM work. Other authors have been recruited from those actively concerned with ECM since World War II as employees of contractors of the Department of Defense, or as members of the staffs of government laboratories. They were selected for their knowledge of particular subjects. The work of writing and publishing this textbook was carried out with the support of the Department of Defense, including all branches of the Armed Services, under U. S. Army Signal Corps sponsorship. It is hoped that this volume will be useful as a textbook for countermeasures training purposes, and as a general reference source for the techniques described, whether or not confined to countermeasures applications.

F. E. Terman

Stanford University
October 1961

x

ACKNOWLEDGMENTS

This book was conceived and initiated by the Department of Defense Technical Advisory Panel on Electronic Countermeasures (TAPEC) under the chairmanship of Dr. Guy Suits. It is the result of advice, counsel, and hard work of many persons, and the cooperation of a number of industrial research organizations, educational institutions, and governmental agencies.

The editors express their deep appreciation to all of those persons and organizations who have contributed to this undertaking.

CONTENTS

xiii

Part I

Introduction

Introduction and Summary

J. A. BOYD, D. B. HARRIS, D. D. KING,
H. W. WELCH, JR.

1.1 Introduction

The first large-scale introduction of electronics into military operations took place during World War II. Since that time, the part played by electronics in weapons systems has increased, often to the point of dominence. This growth in military electronics has been characterized by a profusion of diverse techniques aimed at fulfilling particular military requirements.

Probably the most confusing and little understood aspect of military electronics deals with countermeasures. Since it is concerned exclusively with other electronic devices, primarily in the possession of the enemy, electronic countermeasures is removed from the main stream of weapons technology. However, the growing dependence of modern weapons on electronics and a recognition of their vulnerability has increased tremendously the importance of countermeasures. The complex techniques evolved to effectively counter electronically aided weapons have not been surveyed and made available to the practicing engineer since the publication of wartime accomplishments. The present volume provides a summary of modern countermeasures technology for those working in the field. The scientific and engineering aspects, as well as the military requirements, have been treated; both technical and operational problems must be solved to achieve successful countermeasures. Therefore, both engineers engaged in equipment development and those concerned with operational characteristics should find this book valuable.

1.2 Definition

There has been a tendency in recent years to distinguish between electronic countermeasures (ECM), electronic reconnaissance, and electronic counter-countermeasures (ECCM). When taken together these functions are usually referred to as electronic warfare (EW). In practice, one can seldom completely separate these functions. It is necessary to conduct electronic reconnaissance both for direction of the ECM research and development program and for the operational (tactical and strategic) application of ECM. In this book, emphasis is on electronic countermeasures and electronic reconnaissance; however, studies of jamming effectiveness are pertinent to the evaluation of jamming techniques and counter-countermeasures techniques.

Electronic warfare may be defined as the employment of electronic devices and techniques for the purposes of:

(a) Determining the existence and disposition of the enemy's electronic aids to warfare.

(b) Destroying or degrading the effectiveness of the enemy's electronic aids to warfare.

(c) Preventing the destruction of the effectiveness of friendly electronic aids.

1.3 Peculiarities of EW Systems

Electronic warfare systems occupy a special position in that their primary function is to be responsive to enemy action or potential. The character of effective EW systems and their development cycle does not follow the pattern set by other active weapons and electronic systems and subsystems. The salient points of difference may be listed as follows:

(a) The need for EW systems is recognized when the existence of enemy electronic aids has been established or postulated.

(b) The characteristics of EW systems are determined by the nature of enemy electronic devices—known or anticipated.

(c) The effectiveness of an EW system cannot be demonstrated independently of enemy devices, either real or simulated.

(d) The future course of EW can only be predicted in terms of the anticipated electronic environment to be created by the enemy.

The dependence of EW methods on the present and future enemy electronic posture places the entire field of EW in a particularly close relationship with the intelligence community. The techniques of signal intercept, analysis, and location are primary tools for electronic intelligence (ELINT) and communication intelligence (COMINT). Conversely, the information on

enemy activity and its interpretation is basic to EW development and planning. In the case of a complex transmission system, classification of the signal as ELINT or COMINT may not be a simple matter.

1.4   Research

The development of electronic countermeasures systems places unusual demands on techniques and components research in that:

(a) Operational requirements are continually changing with the development by the enemy of electronic aids to warfare which are the potential target of countermeasures.

(b) Characteristics and vulnerability of target systems are known only through tests made with the aid of countermeasures and reconnaissance equipment developed to meet these operational requirements.

(c) The potential utilization of the complete frequency spectrum, all types of modulation, and maximum efficiency and security of information handling in the target system requires extreme versatility in devices and techniques in terms of design and operational parameters.

Frequently these demands require the use of techniques and components which are not fully matured. Much of the research is directed toward evaluations of feasibility and "trade-offs" inherent in the choice from a multiplicity of alternative approaches to a given problem. An intercept or jamming system designed specifically for a given target system is limited in its application to other target systems. On the other hand, a system designed to handle a number of target systems is limited in its capability against specific targets and is usually extraordinarily complex from an operational standpoint.

Certain techniques, such as the ability to sort and analyze signals and selectively radiate large amounts of power over a wide frequency range, have relatively little value for applications other than countermeasures. The development of these techniques requires in many cases the development of components such as electronically tunable devices, broadband amplifiers and mixers, high power CW oscillators, and noise generators, which in turn find little application in other than countermeasures equipment. On the other hand, since the research on these techniques and components is continuously pushing the state of the art, much of our knowledge of the limits of electronic performance has resulted from the countermeasures effort.

1.5   Summary of Subjects Treated

In the following sections, a short summary is given of each subject treated in this book, with emphasis on the particularly significant results presented in each case.

1.5.1   Introduction and Summary

This chapter is adequately represented by the material in Sections 1.1 through 1.4 and needs no further summary.

1.5.2   History of Electronic Countermeasures

This chapter gives the historical background of the ECM research and development program and discusses features of this program which are unique. The development of ECM since World War II is traced.

Techniques developed in the World War II programs are discussed, and the general characteristics of specific equipments are described. Some of these equipments are now obsolete; some are still in use. Emphasis is laid on the evolution of ECM techniques which led to modern requirements and methods.

A section of Chapter 2 deals with the postwar research and development program in the ECM field. Many of the equipments described here are in current use. The reasons for developing the techniques described and their effect on the boundaries and objectives of the ECM research and development program now in progress are discussed.

No attempt is made to provide a complete list of equipment, techniques, specifications, and applications. Instead, emphasis is laid on the problems encountered and the ways in which they are solved. The program carried out immediately after the war for appraising the effectiveness of electronic warfare techniques, particularly as applied against the Germans, is described. Information was obtained on the spot, by observers in Germany, as to the experiences of German radar personnel operating their equipment while it was being jammed by Allied airborne ECM equipment. This information led to the conclusion that the ECM program was highly successful, and greatly reduced the effectiveness of the German radar.

1.5.3   Perspective of ECM in Modern Warfare

The ways in which the military situation and the geography of the battle and supporting areas affect the problem of destroying or degrading the utility of enemy communications, weapon systems, and surveillance devices is considered in detail in Chapter 3.

Airborne ECM has both defensive and reconnaissance functions; each is considered separately as well as in systems that combine them. Chapter 3 also includes consideration of countermeasures against early warning radar, airborne intercept, and tracking radars, and communications, guidance, and fuze countermeasures.

Utilization of ECM by naval forces is discussed.

ECM in air defense is largely concerned with countering bombing and navigation radar. Passive detection by scanning receivers and target track-

ing also have an important place in air defense; several types of systems are illustrated.

ECM in ground operations is considered from the tactical employment standpoint only. Some of these applications are: ECM against surveillance drone systems, against unenciphered tactical communication nets, against enemy mortar and artillery tracking radars, against tactical bombing radars, against missile systems, and against electronic surveillance devices such as intercept receivers, MTI radars, and infrared scanners, both ground-based and airborne. Predetonation of variable-time fuses, and counter-counter-measures against ECM repeater-jammers are briefly considered.

ECM in aero-space has not yet been thoroughly investigated. Two areas of interest are countermeasures against electronic surveillance of objects on earth, and against AICBM complexes.

1.5.4   Operational Objectives of Intercept Systems

Intercept systems gather reconnaissance information by receiving and analyzing enemy signals. The knowledge gained by this operation permits a more accurate assessment of enemy facilities and preparations than would otherwise be possible. The general value of intercept systems lies in providing information on the enemy signal environment which is useful from an intelligence point of view. The location and character of enemy electronic emitters such as radars, navigation aids, and communications transmitters clearly has a direct operational significance. The application of such intercept data to active electronic countermeasures represents a specific tactical application. Thus, exact knowledge of radar characteristics permits the preparation of optimum electronic countermeasures for use against the particular target radar. In a broader sense, the increasing use of electromagnetic signals in military operations has equally increased the scope and importance of electronic intercept systems. To obtain comprehensive information on enemy radiations is a tremendous task, but the intelligence to be gained on enemy military and technological posture is proportionate.

1.5.5   Signal Environment Study

The design of intercept equipment for the detection, location, and recognition of signals associated with particular radiating equipments depends to a significant extent on the environment in which the intercept receiver will operate. Signal density is an important parameter of this environment. Various workers have approached the problem of predicting signal density in diverse ways. Geographic maps showing typical deployments of tactical units in the field, with their associated radiating equipment, have been prepared, together with complete lists of enemy and friendly radiating

equipment for the particular model chosen. From this information, predicted signal densities have been derived. Studies along these lines lead to a prediction of the number of signals which an intercept system will be required to handle simultaneously.

Recent work includes three signal density studies employing, respectively, a Monte Carlo technique, a deterministic model, and a stochastic-process model. All of these methods give high values for signal density, posing a serious problem for receiver designers. Preliminary studies aimed at comparing the results of the Monte Carlo model and the stochastic process model suggest that perhaps the Monte Carlo method counts the signals from one transmitter too often. In the other approach, the basic assumption of a uniformly random distribution of transmitters is actually violated, and thus its result gives too low an estimate of the signal density in some regions.

Studies based on models, such as those described above, may be supplemented for purposes of verification and of intelligence by observations made in the field by means of a drone aircraft equipped with simple intercept equipment. The drone is equipped with three low-sensitivity receivers along with pulse-counting and recording devices. The system counts the number of pulses received and records the results as a function of time. The signal density is read out and plotted on maps of the tactical area. The contours resulting may then be used to show indications of the locations of major supply areas, depots, and rail heads; concentration of troops; and shifts in air defense.

1.5.6   Intercept Probability and Receiver Parameters

The effectiveness of an intercept receiver depends primarily upon the length of time required for the receiver to intercept a signal, and secondarily on the length of time the intercepted signal continues to activate the receiver. Where CW signals are involved, the length of time required for an intercept is dependent upon the receiver tuning rate and the signal-to-noise ratio; for pulsed signals, signals sweeping in direction or frequency, or both, the foregoing factors are pertinent as well as the probability of frequency and bearing coincidence between the transmitter and receiver.

In the first part of Chapter 6, the probability relationships applying to the intercept problem are developed in a general manner. It is shown, for example, that, if a coincidence probability of unity is assumed, a signal in the form of a 0.5 μsec pulse of S watts peak power masked by gaussian noise having a uniform power density of N watts per cps over a bandwidth of 1 Mc requires a signal-to-noise ratio of 13 db to give a detection probability of 90% on a single trial with a false alarm probability of 0.001. This result applies where everything is known about the signal.

Other relationships are developed for the case of M orthogonal signals where the time of occurrence or the frequency of each signal may be unknown. Curves are provided which permit the prediction of detection probabilities as a function of peak power, detectability index, and false alarm rate. An ideal receiver is considered, and it is shown that this receiver, which is prohibitively expensive, can be simplified by the use of a gated matched filter. The coincidence concept is considered in a general manner, and unit coincidence probabilities and the time required for intercept are developed as functions of system parameters, such as the duration of a pulse, the receiver sweep period, the signal pulse-repetition period, the signal spectral bandwidth, the receiver acceptance band width, the transmitting antenna look period, and the transmitting antenna ation period.

Starting with Section 6.4, the coincidence properties of various types of receivers are considered. It is shown that, for receivers sensitive only to the major lobe of the transmitting antenna, if the acceptance band of the receiver is displaced by its own width in one revolution period, an intercept is certain in one scanning period of the receiver; the time required for intercept is, however, unduly long. Where the scanning period is comparable with the revolution period of the transmitting antenna, results become unpredictable. If the scanning period is comparable to the duration of a "look" of the transmitting antenna, as in the case of a rapid-scan receiver, an intercept during the time of the "look" is certain, and the time required for completing the intercept is small. Where the scanning repetition frequency of the receiver is comparable with the PRF of the transmitted signal, prediction is difficult, but satisfactory results are obtained because of the instability of the system. If the scanning period is comparable with the pulse length of the transmitted signal, an intercept is obtained on the first pulse "seen" by the receiver. In all cases, the proportion of the pulses intercepted is equal to the ratio of the receiver acceptance bandwidth to the receiver scanning bandwidth.

1.5.7 Detection and Analysis of Signals

The theory of signal detectability is a specific application of general statistical decision theory. The problem is one of deciding whether a signal was present in the noise and interference or whether only noise and interference were present. It differs from the usual signal-to-noise ratio approach to receiver design by considering the objectives of the receiver first and working backward toward the actual receiver design. On this basis, designs can be obtained which can be said to be truly optimum; and nonoptimum receivers can be rated against the optimum.

The history, mathematical development, and applications of the theory

are reviewed. Multiple-decision problems and sequential detection processes as extensions of the theory are then discussed.

The second section of the chapter summarizes methods of mathematical signal analysis that have been used in the past in various fields, including electromagnetic reconnaissance. In reconnaissance, one of the major problems is that of defining a set of significant measurable properties of the class of all signals. The relevant equations for the decomposition into classes and subclasses of signals are given and analyzed. Some physical measurements and apparatus useful in analyzing signals are discussed; the measurements are meaningful for both stochastic and deterministic signals. A low-order statistical analyzer system for order-of-battle electromagnetic reconnaissance is given, followed by a description of the signal intercepts at the output of a reconnaissance receiver.

1.5.8   Psychophysics in Electronic Warfare

The human being may be considered an integral part of many systems employed in the data collection, data storage, data dissemination and decision-making processes involved in modern warfare The "operating characteristics" of the human observer in man-machine systems may be specified and quantified through the use of psychophysics.

Psychophysics employs the experimental methodology of psychology and makes use of statistical decision theory and information theory in determining the limitations and capabilities of the human component.

Chapter 8 gives examples of analyses of conventional communications systems, radar systems, and the effect of countermeasures action on these systems. The applications of the methods of psychophysics to the countermeasures problem are described in detail.

1.5.9   The Intercept Receiver

Factors affecting intercept receiver design are considered, including lack of *a priori* information; inability to use integration techniques; complexity of signal characteristics; divergences in operational requirements; divergences in physical requirements; requirements for wide frequency ranges and wide dynamic ranges encountered; complex data handling problems; and the presence of false signals. The intercept probability problem is appraised with particular emphasis on the effect of this parameter on receiver design. The necessity, in designing a receiver, for taking into account the high signal densities encountered, and the consequent data handling problems are discussed. Consideration is given to the relationship of an intercept receiver to a complete intercept system which, in addition to the receiver, includes the antenna, possible auxiliary display equipment, data processing and recording equipment, etc.

A very general relationship is introduced, leading to a figure of merit based only on the composite abilities of an intercept system to monitor simultaneously both frequency and volume of space. The capabilities of receivers having various characteristics and configurations is reviewed in the light of this figure of merit. Sensitivity standards are considered, and methods for estimating received power and signal-to-noise ratio are described. The important receiver characteristics, such as noise figure, bandwidth, gain and dynamic range, tuning range and tunability, and spurious signal response are considered, as they are affected by the environment in which the receiver is required to operate, taking into consideration such factors as intercept probability and the time required to achieve interception. Direct detection receivers for broadband operation, with or without rf preamplification or tunable preselection, including wide-open electronically sweeping and multiple-channel detection types are described. Similarly, superheterodyne receivers, mechanically or electronically tuned and with or without rf preamplification, are appraised. A particular case is the microsweep superheterodyne, which has the capability of sweeping through its entire tuning range in the period of one radar pulse, thus achieving perfect intercept receiving probability. Special requirements for CW reception, the interception of "variable frequency" radars, other signal recognition problems, and receiver "look through" are reviewed.

1.5.10   Direction Finding

The location of enemy emitters is a prime intelligence datum to be gained from an intercept system. To determine the source location, the direction of arrival of the received signal must be measured at several points. The intersection of the direction lines then locates the source. The necessary angle coordinates are furnished by the direction finder at each location. Antennas used for this purpose sample the amplitude or phase delay of the incident wave front. Sequential or instantaneous comparisons then indicate the angle of arrival. The accuracy in angle measurement generally increases with the available antenna dimensions in wavelengths. Significant bearing errors are, of course, introduced by propagation effects which may distort the wave in a variety of ways. The techniques chosen for measuring the direction of arrival vary widely, and depend on the type of signal being intercepted, on the operating wavelengths, and on the local environment, i.e., on whether air, sea, or land-based operation is required. Although direction-finding methods occupy an important place in navigation, communication, and detection systems, the reconnaissance function emphasized in Chapter 10 offers perhaps the broadest application of direction-finding techniques.

1.5.11   The Analysis of Reconnaissance Information from the Data Handling Point of View

Electronic reconnaissance data contains information on the location and technical character of the various enemy emitters. The analysis of the data collected aims to extract this information. Depending on the degree of precision reached, the results are labeled Electronic Order of Battle (EOB) or Electronic Intelligence (ELINT). The types of receivers available for collecting enemy signals select a certain sample of the enemy emissions; the nature of the sample depends on the receiver characteristics, but generally it contains a tremendous mass of largely redundant data. Analysis of this mass of data involves the perception of previously established forms. The primary analysis function provides rapid classification and indexing of the incoming signals. In many cases, the primary classifications provide an adequate identification of the signals involved. A few of the coarse categories may be further studied in a secondary analysis designed to extract the maximum information from a given signal. Many of the operations required in analysis such as digitalizing, sorting, and storing require high-speed machine techniques. Others, such as complex waveform analysis and language translation, demand human assistance. Of course, the final step in analysis, namely, interpretation and dissemination of intelligence information, is always a human operation.

1.5.12   Basic Types of Masking Jammers

The general characteristics of masking jammer transmitters, jammers capable of obscuring information, are treated in Chapter 12. The bandwidth of a jammer and its radiated power are both important factors in masking jamming. Wide bandwidths are desirable if radars and radio receivers covering a broad frequency spectrum are to be jammed. High output powers are required to jam high-powered radars or to jam at long ranges. Generally, bandwidth may be traded for power; i.e., high jammer power output may be achieved by sacrificing bandwidth. Masking jamming is also affected greatly by the type of modulation used. The modulation type, such as amplitude or frequency modulation, and the modulating waveform, such as a sine-wave or sawtooth function, determine how the available jamming energy is distributed in the frequency spectrum. Usually, it is desirable to distribute the jamming energy evenly over the bandwidth with the signal amplitudes following a gaussian distribution. The spectrum obtained from frequency modulation by noise combined with frequency modulation by a sine wave approaches the "ideal" spectrum. Another means of obtaining such a spectrum is the direct amplification of noise (DINA). In practice, amplitude distribution is not truly gaussian because high noise peaks are sliced off either in-

tentionally or because of equipment limitations. Clipping can reduce peak power requirements and can minimize power wastage in the carrier.

Masking jammers may be placed in the categories of barrage jammers, spot jammers, swept jammers, and sweep-lock jammers. Some jammers fall into two or more of these categories by having alternative modes of operation. Barrage jammers are wideband noise transmitters designed to prevent use of frequencies over wide portions of the electromagnetic spectrum. Many receivers can be jammed simultaneously, or frequency-diversity radars can be jammed without readjusting the jamming frequency. Barrage jammers may not work effectively against systems which use high-powered transmitters because jammer power may not be sufficiently high in the transmitter frequency band. Spot jammers, manually tunable transmitters which are amplitude or frequency modulated by noise or by a periodic function, can concentrate high power in a narrow band; these jammers can be used to mask specific transmitters from communications or radar receivers. Where tunable receivers cannot be used as an antijam feature, spot jammers can be used to good advantage. Swept jammers are transmitters in which a narrowband jamming signal is tuned over a broad frequency band. These jammers combine the high power capabilities of spot jammers and the broad bandwidth of barrage jammers. An important factor in considering the effectiveness of swept jammers is the dwell time, the period during which jammer noise is in a receiver's bandpass. A sweep lock-on jammer is essentially a swept jammer with the additional feature of lock-on capability. The sweep lock-on jammer can concentrate much noise power in a narrow band. This type of jammer can lock on to a second signal much more quickly than a spot jammer can.

1.5.18   Geometry of the Jamming Problem

The locations of jammers and receivers affect the required jamming power levels. Equations are readily derived for particular geometries and particular jamming situations; parameters entering into these equations must be carefully considered for each individual problem. As would be expected, the threshold of intelligibility must be defined differently for each system and for each type of jamming signal. Some of the factors which appear when considering geometrical relations in radar jamming include the jamming noise-to-signal ratio at the receiver, receiver bandwidth, range between jammer and receivers, antenna gains, and radar cross sections. How a particular jamming situation is examined determines which of these and other factors appear in the equations. For example, if a self-screening airborne radar jammer is aboard the target aircraft, the minimum range at which the jammer is effective is expressed in terms of the radar peak power, the radar

transmitter antenna gain, the radar receiver bandpass, the jammer power per megacycle, the jammer antenna gain, the radar cross section of the target, the ground reflection factor, and the camouflage factor. For the case of jamming a search radar from a target aircraft carrying the jammer, many of these same expressions can be used in deriving an equation. Here the final equation may be a comparison only of maximum detection ranges out of and in a jamming environment. The ratio of the ranges can be expressed simply in terms of the normal minimum detectable signal and the minimum detectable signal in the presence of jamming noise. For communications jamming, two ranges, that between the communications transmitter and the receiver and that between the jammer transmitter and receiver, must be considered. In a case such as this one, more attention must be paid to propagation conditions since the transmission paths are not the same for the communications and jamming signals. The geometrical picture of the radio proximity fuze jamming problem is more complex than those of communications and radar problems. It is necessary to consider the trajectory of the fuzed missile, the location of the jammer, and the target area to be protected. If a fuzed missile is to be predetonated at a certain height above the ground, the jamming field strength must be sufficiently greater than the threshold field strength above this height. Other, more complex, geometry problems involving the use of jammers exist. For example, a difficult situation to evaluate would be one in which multiple jammers are used against air-to-air missiles by aircraft flying in formation. In such a case, to determine the volume of protection, the pattern of the jammers' antennas, the jammer powers, the spacing between jammers, and the direction of arrival of a missile, all must be taken into account.

1.5.14   Effectiveness of Jamming Signals

The problems of evaluating the effectiveness of jamming signals in search radar, tracking radar, and communications situations are treated in Chapter 14 with consideration being given to those signals established as being feasible of generation and effective as electronic countermeasures.

First, the general concepts of determination of effectiveness of signals is considered; then search radar is discussed from both theoretical and experimental points of view. Criteria are established, and the responses of receivers to DINA and FM-by-noise jamming are ascertained. The influence of standard antijamming techniques on the effectiveness of the jamming is considered.

In tracking radar, the influence of the type of tracking employed on jamming signals to be used is established. Experimental methods of determining jamming effectiveness are discussed.

The jamming of data links is considered only briefly. Results of jamming

voice-communications systems of common types where continued degradation rather than threshold type of failure is experienced are shown to be dependent on the evasive action permitted the receiver, particularly in the FM case. A valid measurement technique for the establishment of jamming effectiveness in communications systems is noted.

At the time of the writing of Chapter 14, little quantitative information was available regarding the effectiveness of signals in jamming of single-sideband and pseudorandom communication links.

1.5.15   Radar ECM Repeaters and Transponders

The advent of broadband radio-frequency amplifiers, such as traveling-wave tubes and distributed amplifiers, and of electronically tunable oscillators, such as backward-wave oscillators or carcinotrons, has made possible the development of jamming systems which are relatively sophisticated in comparison with brute-force barrage-jamming equipment. These sophisticated systems generally are designed to deceive the radar operator by producing, on the radar scope, false targets or targets giving incorrect information, as contrasted with barrage or spot-jamming systems, which depend upon obscuration of the radar scope. Deception-type systems may employ either repeaters (devices which amplify and reradiate the radar signal, adding incorrect information) or transponders (devices which produce similar effects by means of local oscillators and memory circuits). Repeaters and transponders have advantages with respect to brute-force devices in that their power-output requirements are relatively low; and in many situations, deception is achieved without betraying the fact that countermeasures are being employed. On the other hand, barrage and spot jammers possess the advantage of universality.

Analysis shows that the gain required in a self-screening repeater is independent of the range, and is, in fact, a function of the ratio of the effective radar cross section to be simulated to the capture area of an isotropic radiator. On the other hand, the power requirement is in accord with the usual jamming equation, and is inversely proportional to the square of the range.

Various types of repeaters are described, including straight-through repeaters, swept repeaters, and gated repeaters. Such repeaters may be used to change the apparent size of the target. In the case of the swept repeater, the disruption of the operation of semi-active systems utilizing doppler shift may be effected.

In addition, as discussed in later sections of Chapter 15, various deception techniques can be incorporated to produce false targets or to generate false range or bearing information.

Single frequency transponders have been used for some time to produce

false target information in a limited frequency range. This technique has been extended to the development of search-lock-jam transponders covering a broad band of frequencies, utilizing traveling-wave tubes and backward-wave oscillators. In such transponders, a sweeping oscillator stops sweeping when it encounters the radar signal and triggers on a jamming transmitter on the same frequency. Such search-lock-jam transponders scan the entire frequency band in the period of a single pulse.

Devices for obtaining broadband amplification are discussed, including distributed amplifiers for low frequencies (below 400 Mc) and traveling-wave tubes for microwave frequencies. Electronically tunable resonant circuits may be used for the rapid tuning of oscillators in the low frequency ranges. Backward-wave oscillators and carcinotrons are useful in the microwave range.

Deception techniques include false target generation, employing straight-through repeaters sensitive to minor lobes of the radar antenna, and generating false targets of a constant range but varying azimuth. Another highly effective deception technique, employed against tracking radar, is "range gate pull-off." This technique is implemented by means of a repeater with frequency memory, which records the frequency of the received signal and reradiates the signal after a lapse of time, which is successively and continuously increased in order to pull off the range gate in the radar receiving equipment. Other repeaters or transponders may employ scan-rate modulation to generate large error signals in the radar servo loop. Bearing errors may be introduced into lobe switching or conical scan radar tracking systems by means of inverse gain modulation, which returns strong signals when weak ones are received, and vice versa. The velocity gate in active radar systems of the CW doppler type may also be pulled off by varying the frequency of the return of the repeater or transponder.

1.5.16 Fuze and Communications Repeaters

The application of repeater-type jammers against radio doppler proximity fuzes and communication links is discussed. In comparison with other fuze countermeasures equipments, repeaters have a high potential against broad classes of radio doppler fuzes. Simple repeater jammers have little merit against communication links; however, variations from a simple repeater may have value as a communication jammer.

A brief theoretical discussion is given of the single-channel radio doppler fuze, including the following types: CW, FM, PD (pulsed doppler), and noise. A more general description is given for special fuzes and multi-channel fuzes. A basic range equation which relates the range at which a fuze will be predetonated to the fuze parameters and repeater jammer parameter is given.

The problem of obtaining sufficient isolation between input and output of repeaters to prevent oscillation has led to wide use of time-shared repeaters. Distributed amplifiers, when used as the final amplifier of superheterodyne time-shared receivers, also serve as the transmit-receive switch. Basic configurations and special problems associated with superheterodyne time-shared repeaters are discussed. The special characteristics and applications of long- and short-delay repeaters are discussed. The operation of a superregenerative repeater is described, and its application as a fuze jammer is discussed. The usefulness of the superregenerative repeater is limited by the fact that it is impractical to maintain an artificial doppler shift, and the rf band which can be covered effectively with such a jammer is very narrow.

Care must be taken in the use of repeaters against voice-communication links since simple repeaters may actually augment the direct transmission. A "babble-of-voices" jamming signal can be generated by demodulating the communication signal then retransmitting the carrier with some new form of modulation which may be related to the original modulation.

Coded radio-transmission links are classified as asynchronous or synchronous. The information bandwidth of asynchronous coded radio links, such as common teletype, is small, and the spectrum of the transmitted signal is comparable in bandwidth. Efficient jammers for this type link can be produced. As the degree of synchronism is increased, the jamming problem can be made increasingly difficult. Special features of synchronous systems using pseudo-noise encoding, which make them susceptible to jamming, are discussed.

The advantages of jammers with provisions for monitoring while jamming are discussed. Examples of jammers of this type are cited.

1.5.17 Programmed Automatic Jamming Systems

The timeliness which is essential to the effectiveness of many jamming operations has created a continuing interest in automatic systems. For single cases, such as jamming of a specific type fuze, a repeater which responds only to the specific type of signal radiated by the fuze is all that is involved. For more complex cases, such as protection of a strategic bombing mission that passes through friendly, early warning, area defense and local defense zones, more elaborate programming is required. The key to programming is mode selection. Selection of mode of operation may depend on instructions fed into the programmer before the mission, updated instructions communicated to the vehicle carrying the jammer during the mission, or current information based on intercept and analysis of enemy signals during the actual jamming periods. In the example of the strategic bombing missions many inputs and responses must be accounted for. Inputs may be: type of radar to be jammed, direction of radar, frequency, scan rate, pulse repetition fre-

quency, etc. Responses may be designed to deceive the enemy as to the direction or size of the attack force or to conceal the attack force completely. Capacity of the mode selector in terms of numbers of input signals, instructions, and decisions possible must be compatible with reliability and weight requirements as well as with the type of mission. Geographical programming may be used, based either on preprogrammed mission instructions or on inputs from a navigational system. This method may provide weight and simplicity advantages. In all cases adequate attention must be given to interference and electronic compatibility and to provision for self-testing of components or programs within the jammer.

1.5.18 Confusion Reflectors

The radar echo from a target can be masked not only by jamming signals but also by other echoes. Such echoes are produced by confusion reflectors specifically designed for this purpose. The simplest confusion reflectors, called chaff, are very fine foil strips cut to resonant lengths. At longer wavelengths, untuned lengths of foil, called rope, are commonly used. In both cases, the echoing area is very large in proportion to the weight of material. Indeed, the basic purpose of confusion reflectors is to provide a large echo over a broad frequency band in a compact, lightweight package. The designer of confusion reflectors attempts to place a sufficient number of false targets about the real target to render difficult its radar identification and tracking. This requires rapid and uniform dispersal of the individual elements after ejection, as well as a slow rate of fall of the resulting cloud. In principle, the different velocity of the real target permits the radar to discriminate against surrounding confusion reflectors. However, the techniques required to perform the necessary discrimination limit the over-all performance of the radar. Only a slight degradation of this sort may strongly affect the outcome of an engagement. For example, a tracking radar may be forced to reacquire its target because of the noiselike signal introduced by chaff. In an engagement between high-velocity vehicles, the resulting loss of guidance information may suffice to cause the attack to fail. Similarly, in a search radar system, when the number of targets approaches the available capacity, the introduction of added discrimination requirements may saturate the system.

1.5.19 Target Masking and Modification

The echo from a target is determined by its geometrical shape and by the reflectivity of its surface. By controlling these factors, the radar visibility of a target can be reduced or increased. In the case of real targets, a reduction is desired, while in the case of false targets, enhancement of the echo

is sought. Shape factors control the directional properties of the reflected signal. Thus, a flat surface, large compared to the wavelength, concentrates the reflected energy in a narrow beam; unless the surface is exactly normal to the incident signal, there is no reflection toward the source. Conversely, three reflecting planes at right angles reflect the incident energy back toward the source. These two types of surfaces respectively reduce and enhance the radar visibility over a wide range of aspect angles. Absorbent coatings can be made effective over a limited band of frequencies and angles of incidence. In general, the wider the range of performance, the bulkier the material becomes. The extent to which the countermeasures designer can control shape and surface parameters is usually limited. This is particularly true in the case of airframes. However, for ground targets, very effective transformations of the radar scenery have been achieved.

1.5.20   Decoys

An effective defense must send weapons against all targets that are threats. When the number of threatening targets exceeds the supply of weapons, then the excess cannot be engaged at all. The aim of decoys is to provide enough false targets interspersed with the real ones to saturate the defense in this manner. The key to the decoy problem is evidently the discrimination between true and false targets which can be exercised by the defense within the time and space limits of the engagement. Accordingly, the decoy designer attempts to make discrimination difficult by closely simulating the radar and infrared characteristics of the real target. Since the decoy must be substantially less costly than the real target, only the critical characteristics can be treated. Even with this limitation, the cost of a flock of decoys expended in one engagement may be high. Against this cost must be balanced the cost of other protective measures. The possibility of improvements in radar discrimination must also be considered, since a single change in radar technique might unmask the decoys and render them useless. The interplay of these factors gives the decoy problem broad scope in countermeasures, even though its practical application has been limited. To illustrate the many factors involved, the problem of decoys for heavy bombers is considered in Chapter 20. The ground defense, bomber, and decoy characteristics all enter this study.

1.5.21   Characteristics of Infrared Radiation

Infrared refers to the portion of the electromagnetic spectrum lying between the visible and short-wavelength microwaves. Since all objects at a temperature above absolute zero emit infrared radiation to some extent, many objects of military importance are unavoidably good infrared targets.

The wavelength region from 1.2 to 7 microns is at present the most important from the military standpoint.

Some of the advantages of infrared over microwaves follow. (1) Much greater resolution can be obtained with infrared devices, which use very much smaller "apertures" than microwave devices do. (2) Many infrared systems are passive. (3) In an active system (e.g., communications), the energy cannot be intercepted by th. enemy without getting directly in the beam. (4) The electronic circuits are usually simpler and cheaper than those in microwave devices.

Some disadvantages of infrared as compared to microwaves follow. (1) Clouds and water vapor greatly reduce the effectiveness of infrared systems. (2) Background radiation, particularly in daylight, is often troublesome. (3) Infrared sources and detectors cannot be "tuned" as sharply as microwave devices. (4) Passive infrared systems do not give range information.

Detectors which have found the widest application among military infrared devices are the high-sensitivity, short-time-constant photoconductors such as lead sulfide, lead telluride, and indium antimonide. Thermal detectors such as bolometers and thermopiles are used less frequently in military applications, especially in the short-wavelength regions, because of their relatively slow response and, frequently, low sensitivity.

The chief military uses of infrared are in tracking, in mapping, and in communications. Tracking devices have found wide application in air-to-air and surface-to-air missile systems. These passive devices track the radiation from aircraft engines. A number of infrared scanning devices have also been developed for obtaining high-resolution maps of ground inst :tions. Infrared strip maps can be used to identify objects which cannot be identified from photographs.

1.5.22 Infrared Countermeasures Techniques

Most military infrared devices are passive in that no active source is required to illuminate a potential target. These devices operate solely on radiation which originates from their intended target. Because of this, practically all infrared countermeasures are based on two basic concepts; these are the concept of a false target and the concept of suppressing radiation from potential targets. Considerable study has been devoted to the use of decoys as false targets against air-to-air missiles attacking subsonic jet aircraft. The decoy in its most common form consists of a pyrotechnic flare similar to those which have been used for illumination and identification purposes. Those factors which are important in designing and using a flare properly include infrared intensity ratio between flare and aircraft, flare trajectory with respect to the flare-dispensing aircraft, characteristics of the missile

seeker, and the missile guidance criterion. Decoys might be used more effectively if they are used in conjunction with some form of radiation reduction. Shielding and cooling of jet engines can be used to alter the radiation pattern of aircraft such that the zone in which these aircraft can be tracked by infrared seekers is reduced. Infrared engagement warning systems can be used to detect rocket radiation of air-to-air missiles. When a launch is detected, countermeasures can be initiated. Such warning systems do not have the capability of indicating whether the missile is an infrared or radar-guided weapon.

Decoy radiation sources and camouflage techniques could conceivably be used to protect ground installations. However, the most effective means for the protection of ground installations appears to be the use of screening agents such as smokes composed of plastic particles. Principal problems include producing a sufficient quantity of smoke to protect a large area, the settling rate of the particles, and the toxicity of the material used.

1.5.23 Underwater Acoustic Countermeasures

Chapter 23 describes the basic concepts of underwater acoustic countermeasures, including system description and design requirements. The ocean presents a very poor environment for the propagation of electromagnetic waves; therefore, underwater communication and detection must rely on other forms of energy propagation. Acoustic energy is most commonly used, both in the sonic and ultrasonic frequency ranges. The word "sonar" is used to describe, in general, the detection and tracking of underwater targets, whether by passive listening or actively transmitting pulses.

The general functions of intercept, jamming or masking, and deception are discussed, and the special problems associated with surface ship and submarines are considered. Nonacoustic methods of detection are currently unable to detect a submerged submarine, although research efforts which may prove effective are being made. The modern submarine is essentially a true submersible; thus, acoustics provide the only ready approach to detecting and destroying submarines. Submarines, therefore, should be primarily capable of countering acoustic detection and attack.

The submarine platform is used as the basis for developing a countermeasure system and showing the relationships of the subsystems. A functionally integrated system is described which performs the functions of intercept, signal analysis and display, tracking of noisy targets, and automatic control of both mounted and expendable jammers.

Design requirements for interception, jamming, deception, and vehicles are given. The basic difference between an intercept receiver and a passive sonar is that the former is designed to intercept deliberately transmitted

signals, whereas the latter is designed to detect noise ships make in transit-
ing the seas. Special problems associated with the design of passive sonar
systems are discussed. Requirements based on system operating parameters
are given.

The relative merits of masking and jamming are given. Design require-
ments for jammers are discussed in relation to the system-operating para-
meters. Specific jammers are described and examples of jamming effective-
ness against sonar are given.

The general battle situation in which the need for deception devices arises
is described, and decoys which may be used to simulate a submarine are
discussed. Jammers used in protecting either submarines or surface ships
may be mounted on towed or expendable-type vehicles. The requirements
for such vehicles are considerably different for the two platforms. The re-
quirements and design of these vehicles are discussed.

1.5.24   Circuits

Chapter 24 is a compendium on circuit techniques, techniques which may
be applied to devices in many fields besides that of countermeasures. Linear
video amplifiers are described with emphasis on the time or transient re-
sponse of these amplifiers to a square-wave input. How rise times, consider-
ably shorter than ... se obtainable with video amplifiers, can be obtained
with additive amplifiers (such as distributed amplifiers and split-band ampli-
fiers) is also discussed.

A section on filter amplifiers treats the subject from the standpoint of
frequency response. In many systems, there is a requirement for producing
a signal gain over a band and for rejection of signals outside of this band.
As examples, i-f amplifiers, with and without staggered tuning, are discussed.
The effects of single and double-tuned interstage networks are also described.

The general characteristics of passive frequency-selective filters are de-
scribed briefly. There is a similar section on time-domain or correlation
filters. Other circuits briefly discussed include amplitude limiters, super-
regenerative circuits, and locking-in oscillators.

1.5.25   Mechanically Tuned High-Power Oscillators and Ampli-
fiers

Historically the development of new applications of electronics in war-
fare, other than countermeasures, has gone hand in hand with the develop-
ment of new devices for power generation, modulation, tuning, rf switching,
etc. For example, radar development was made possible and has continued
with development of high-power, pulse-modulated tubes; high-power modu-
lator tubes; gaseous TR tubes; ferrite isolators circulators; and antenna

switching elements. Requirements and the value of countermeasures equipment to destroy or minimize the effectiveness of radar were definable as radar characteristics; and limitations become known. Thus, the general characteristics of devices useful in the countermeasures application have been defined and are well known. Some of these are: broad bandwidth, high cw power, wide tuning range, rapid tuning, high efficiency, and susceptiblity to a variety of modulations. It is characteristic of countermeasures-device developments that they have been "forced" in reaction to the development of target equipment. This has led to much use of state of the art development and relatively rapid obsolescence.

Chapters 25 through 28 cover the background and present picture of device development.

The first significant series of developments beginning during World War II and centered around more or less conventional mechanically tuned triodes, tetrodes, magnetrons, and klystrons. The magnetron by virtue of its basic high efficiency and high power capability has been and still is widely used in the microwave region. Below 1000 Mc, tetrodes have been a strong competitor. The klystron has suffered primarily because of the interrelated relatively low efficiency, tunability, and bandwidth problems. Accessory equipment such as modulators, filament control, focusing magnets, and magnetic supplies have been the source of many difficult problems primarily associated with size and weight of equipment.

Development of counter-countermeasures such as frequency-diversity radar has limited the applicability of mechanically tuned devices and stimulated the development of direct generation of noise for barrage-type jamming and electronically tuned devices discussed in the following sections.

1.5.26 O-Type Microwave Tubes

The invention of the traveling-wave-tube amplifier with its basic broadband characteristics was the beginning of a series of developments capitalizing on the basic principles of interaction with traveling waves in both nonre-entrant and re-entrant structures. Design of beam-forming electron guns, magnetic focusing, and slow-wave structures have progressed together to improve noise characteristics for receiving applications and power and efficiency charac istics for transmitting applications.

The discovery d application of backward-wave interaction led to another important feature, electronic tuning, orders of magnitude faster than was possible mechanically.

This type of tube, classified the O-type tube, includes all traveling-wave types, and amplifiers and oscillators not having a magnetic field perpendicular to the flow of electrons. O-type tubes find their most significant ap-

plication in receivers. As power-transmitting tubes they suffer from low efficiency, generally less than 30% for forward-wave structures and of the order of 10% for backward-wave electronically tunable structures. Backward-wave oscillators have generated frequencies up to 100,000 Mc at milliwatt power levels. Several hundred watts have been achieved at S-band.

1.5.27   Crossed-Field Microwave Tubes

Probably the most significant single factor since World War II in the development of potential for jamming capability has been the increase in the versatility of crossed electric- and magnetic-field microwave power generators and amplifiers. The mechanically tuned magnetron which has been basic to most microwave jamming systems is notable for its high efficiency and relatively high power capabilities. However, it does not provide all that the designer requires in flexibility when one considers modulation, tunability, bandwidth and amplification requirements. The so-called M-type backward-wave oscillators and amplifiers (M for magnetic), Bitermitrons, voltage-tunable magnetrons and, more recently, crossed-field forward-wave amplifiers have done much to remove these limitations.

With the high-power backward-wave oscillators, 200 to 400 watts can be obtained at any frequency from 200 Mc up to 11,000 Mc at 20% to 40% efficiency. Tuning ranges of 30% to 50% are possible. Radiated noise bandwidths up to 25% of the operating frequency are possible. Power outputs in excess of 1000 watts are available below 3000 Mc; 10 kw has been obtained in the 200- to 400-Mc ranges.

With a Bitermitron (backward-wave amplifier) and oscillator driver, combination power outputs in excess of 1000 watts have been obtained in the 2500- to 3300-Mc range, and 450 watts have been obtained at X-band. At high power levels, the Bitermitron is more precisely described as a locked oscillator than as an amplifier. At low power levels, with the tube acting as a superregenerative amplifier, gains from 50 to 70 db have been obtained at S-band. Stability in this mode of operation requires careful attention to the design of accessory circuitry.

The voltage-tunable magnetron has simplicity as its great virtue. It is inherently low power and low efficiency for tuning ranges similar to those of backward-wave oscillators but has many applications as a local oscillator or driver in ECM systems. When it is associated with a power traveling-wave tube, very flexible jamming-system design is possible since it can provide driving powers of 1 to 2 watts tuned over a 2:1 frequency range.

If the tuning range is restricted to 10-15%, 50 to 100 watts of output power is achievable at greater than 50% efficiency. Above 6000 Mc, the power output is restricted to the milliwatt level in the present state of the art.

Future developments promise higher powers and slightly increased efficiencies for all of these tubes combined with the attractive features of wide bandwidth and electronic tuning range.

1.5.28 Ferrimagnetic, Gaseous Electronic, and Ferroelectric Devices

Ferroelectric, ferrite, and gaseous electronic devices seem to be unrelated. However, at frequencies below 500 to 1000 Mc, ferroelectric and ferrite devices exhibit analogous dielectric and magnetic nonlinearities applicable to electronic tuning, modulation, and other parametric applications. At frequencies above these values, ferrite and gaseous electronic devices in the presence of a magnetic field exhibit analogous magnetic and dielectric tensors, dependent on the applied magnetic field, leading to nonreciprocal wave propagation, variable phase shift, wave switching, variable attenuation, and dielectric breakdown. These phenomena find many potential applications in countermeasures equipment.

The control of properties of ferroelectric and ferrite ceramics, as well as gaseous electronic media, is relatively difficult to achieve. Certain inherent properties, such as temperature and frequency sensitivity, also create problems. However, much progress has been made in the application of these devices to tuning of search receivers, antenna switching, modulation, broadband isolation, amplitude regulation, and limiting phase-shift control and polarization rotation. Continuing effort will mature these applications and discover more.

1.5.29 Antennas

The impedance and radiation properties of antennas are considered in general; and in detail for the countermeasures application. The input impedance of the antenna is expressed in terms of the voltage standing-wave ratio. The radiation pattern is a two-dimensional plot of the variation of antenna response as the antenna is rotated about a specified axis. The power gain of an antenna is defined as the ratio of the power delivered by the antenna to the power which would be delivered by an isotropic antenna located at the same point in the incident field. (An isotropic antenna is an idealized, and fictitious, antenna whose radiation pattern is a circle for any orientation of the axis of the pattern.) The directive gain of an antenna is defined in the same fashion as the power gain except that the antenna is assumed to be free of losses. It is useful to refer to the aperture, sometimes called the collecting aperture, or receiving cross section of an antenna, and to define effective gain and effective aperture as the values observed when the load is mismatched by some known amount. Aperture efficiency is the ratio of the aperture of an antenna to its physical cross section.

Bandwidth considerations particularly important for ECM applications are discussed. It is shown that every radiating device is a high-pass structure and that, the higher the Q of an antenna, the more difficult it is to match over a wide band. The jamming equation is stated, and the effect of the gain parameters in this equation is discussed with particular reference to the development of optimum jamming antenna patterns. Examples are given of optimum patterns for various situations. Similar treatment is given to optimum intercept antenna patterns.

Typical ECM antenna designs are described, including electric dipoles, magnetic dipoles, helical and spiral antennas, logarithmically periodic antennas, horn antennas, reflector-type antennas, surface-wave antennas, and antennas for direction finding; and their VSWR and pattern characteristics are discussed. Siting problems are reviewed in a general fashion.

1.5.30   Supplementary Circuits and Techniques

In Chapter 30, an associated group of techniques, which are of special value in the design of countermeasures equipment, are described. The main subheadings of the chapter are Receiver Circuits, Analyzer Circuits, Transmitter Circuits, and Recording Techniques.

Under each subheading a brief discussion is given of special requirements placed on countermeasure equipments, and specific techniques for handling some of the requirements are described. Circuit diagrams or block diagrams are shown for all techniques discussed. All discussions are based on concepts and techniques which have been demonstrated with working models of equipments.

1.5.31   Propagation

Aspects of propagation which are of most direct concern in electronic countermeasures, and particularly in radar countermeasures, are considered in Chapter 31. Since the propagation characteristics of the transmission path affect both active countermeasures (power jamming and deceptive jamming), and passive countermeasures (detection, location, and detailed signal analysis), attention is given to both of these functions, in addition to the propagation question in general.

In the case of line-of-sight propagation, the free-space propagation equation, given in Chapter 31, is affected by molecular absorption, ionospheric dispersion, tropospheric refraction, ionospheric absorption, tropospheric dispersion, ionospheric refraction, and polarization. In the absence of any of these effects, a sample figure of 157-db attenuation applies to transmission at a frequency of 100,000 Mc over a distance of over 10 miles; at 10,000 Mc, over a distance of over 100 miles; at 1000 Mc, over a distance of 1000 miles, or at 100 Mc, over a distance of 10,000 miles. Thus, other things

being equal, free-space propagation favors the lower frequencies. Molecular absorption modifies this simple result considerably. The principal gaseous molecular absorption bands exist at frequencies of 22 kMc (H_2O), 60 kMc (O_2), 119 kMc (O_2) and 170 kMc (H_2O). Because of such molecular absorption, the losses in the transmission path may be increased by unexpectedly large values. For example, vertical transmission to a height of 10 miles at 60 kMc may be subject to an additional 100 db of loss due to oxygen absorption. A tangent ray traveling to a height of 300 miles may experience absorption as high as 4600 db at 60 kMc.

Ionospheric absorption in the case of a radio wave propagating through the ionosphere has a negligible effect at frequencies above 30 Mc. Tropospheric dispersion also has a small effect, in comparison with the absorption suffered over a given path. Phase distortion is encountered due to ionospheric dispersion. However, it is found that relatively narrow pulse widths can be transmitted through the ionosphere without excessive distortion. For example, the minimum pulse width at 1000 Mc is 0.2 μsec; and at 5000 Mc, 0.02 μsec. Accordingly, ionospheric dispersion is not usually a serious problem.

In line-of-sight propagation, variations in the index of refraction of the atmosphere cause small changes in direction of arrival, which, however, may result in destructive interference. Tropospheric refraction, where strong stratification of the atmosphere exists (ducting), results in strong downward bending of rays within line of sight and angles of arrival higher than would be the case under standard conditions. Similar downward bending is caused by ionospheric refraction, even when the frequency is sufficiently high to penetrate the ionosphere. Changes in polarization, when transmitting through the ionosphere, may also affect the line-of-sight case.

Transhorizon propagation may result from the phenomenon of refractive bending. In addition, there are other mechanisms such as diffraction, scattering, and reflections from meteor trails and auroras which help circumvent the limitations of the horizon. Tropospheric refraction resulting from temperature inversions and strong vertical humidity gradients may permit transmission far beyond the horizon. Signal levels in a duct may average around the free-space level appropriate to the distance covered when well beyond the horizon. Ionospheric refraction and reflection are effective at horizon distances generally below 100 Mc. Tropospheric scattering effects result in extended transmission beyond the horizon. A useful, though not exact, rule of thumb for estimating signal levels at frequencies in the general region of several hundred or a few thousand megacycles is to consider the signals at 100 miles to be 60 db below the free-space value, and to be decreasing at 0.17 db per mile. Tropospheric scattering also has the effect of spreading the power over a range of angles $1°$ or $2°$ in extent.

2

History of Electronic Countermeasures

D. B. HARRIS, H. O. LORENZEN, S. STIBER

2.1 The Development of Countermeasures Through World War II

2.1.1 The Early Days of Electronic Warfare

Electronic warfare was first employed as a preliminary to the Battle of Jutland. In the days preceding 31 May 1916, the Admiral of the Fleet, Sir Henry Jackson, employed evidence of coastal radio direction finders under Admiralty supervision to detect movement of the German fleet. The changes in the apparent directions of arrival of radio signals from the enemy fleet were very slight, but Sir Henry dared to move the opposing British fleet on the basis of this information. An interesting related fact is that the British had the code books recovered when the German light cruiser *Magdeburg* ran aground along the coast of the Russian Black Sea two years earlier; hence the damage inflicted by the British was no secret as the Battle of Jutland developed, since the Admiralty was able to decode the German radio messages.

The British Admiralty and the U. S. Navy later worked together as a scientific team, even before 27 June 1940 when the National Defense Research Committee was formally established in the United States. The Navy, as well as the other Services, has continued to be active in the ECM effort, to this day. During World War II the Navy effort was centered at the Naval Research Laboratory in Washington, where in-house research was carried on by the NRL staff. This Naval Laboratory served to fill the Navy's urgent

fleet requirements until the immediate threat could be neutralized and commercial talent could be brought into play to furnish the production quantities of the equipments needed.

Similarly, effort in the ECM field on the part of the U. S. Army, by the Signal Corps, provided the ground-based and airborne U. S. Army Air Corps equipments. Direction-finding (DF) techniques in the hf band had already been developed to give satisfactory performance prior to World War II. Ground-based and airborne direction finders were developed by the U. S. Army Air Corps. When experiments showed radar to be a practical device, DF techniques were extended upward in frequency, and some preliminary thought was given to methods for neutralizing radar technology.

History of ECM during the Battle of Britain. The forerunner of ECM was the jamming and deceptive tactics of the British against the navigational systems of the Germans early in World War II. In June 1940, the Germans established an extensive series of short-wave stations (200 kc to 900 Mc) in northern France. These stations were beamed over London. An aircraft equipped with a loop antenna could get on any one of these beams and follow it directly over London. This quite successful navigational aid was known as *Lorenz*.

After a great amount of study, the British countered with a system they called *Meaconing*. Combinations of receiver and transmitter, separated by 5 to 10 miles, were established. A British receiver picked up the navigational beams from a German transmitter and sent the signal by land lines to a British transmitter. German planes attempting to get bearings then received signals from the original (German) transmitter and the British transmitter or masking beacon (meacon) and obtained either no bearing or the wrong bearing. On several occasions German planes became completely lost and landed on British air fields.

Then the Germans used two intercommunicating transmitters on the French coast, and while one transmitted dots the other transmitted dashes, on parallel beams. When the plane was on the course between the beams it received a steady tone in its receiver. When off to one side, it received either dots or dashes. The steady tone was 0.3° wide. Using this system, the Germans were able to determine their position over London within 900 yards. This *Knickbein* was called *Headache* by the British, who countered with a system called *Aspirin*, a receiving-transmitter setup working on the same principle as Meaconing. This system simply retransmitted the German beams on to a new location.

About September 1940, the navigational war really got down to business. The Germans initiated the use of *Ruffian*, a propaganda project which operated on 70 Mc. The propaganda was used to conceal a directional purpose,

and it fooled the British for a long time. Shortly before a raid was scheduled, this propaganda beam continued broadcasting but was narrowed down to 5°, and this, called the pilot beam, was the beam that the formations followed over the target. In addition, they had a second 5° beam which crossed the pilot beam at predetermined points. A clocking device, which gave the plane's ground speed, was turned on when the plane crossed the pilot beam. Then, according to his air speed, after the plane crossed the second beam, the bombardier dropped his bombs at certain points or time intervals. The method used in obtaining such a narrow beamwidth has never been determined.

The British counter for *Ruffian* was *Bromide*. Again, the Meaconing idea was used on a different frequency. The British rebroadcast the narrow pilot beam with a nondirectional antenna, making the beam useless for navigation. They also used directional antennas to rebroadcast beams across the pilot beam in such a manner that enemy bomb loads would be dropped in the English Channel. During this period, newspapers carried the story that the bombers dropped their loads in the Channel because they were running from the British Spitfires. This was a security measure taken to keep the Germans from realizing the effectiveness of British countermeasures. *Ruffian* was used by the Germans until January 1941.

The third phase of the navigational war brought out still another complicated German system called *Benito*. The *Benito* consisted of a 45-Mc beam, frequency modulated. This again was the pilot beam with the London area as the target. The lead plane in a formation was equipped with a transmitter. The plane flying down the beam received and retransmitted the modulated signal back to the ground station. By measuring the phase difference between the transmitted signal and the received signal, the ground station could inform the plane by radio of its exact distance from London. After a period during which London was heavily bombed, the British countered with the *Domino*. They again rebroadcast the German beam with a nondirectional antenna. As a variation, the British rebroadcast the German beam so that it would appear as if the planes were far from London, or past London. Another variation which vexed the German pilots very much was a German speaking voice coming in on German receivers, giving them wrong directions. This voice, originating in an English station, often led the German bombers to land at British airports. *Benito* was used by the Germans until about June of 1941.

It seemed that the Germans had exhausted their power to build new navigational aids, but they still managed to devise new methods for using the old systems. The Germans equipped one squadron (KG-100) with all their available navigational aids and used the various types alternately. The KG-

100 dropped incendiaries on London, making it possible for bombing planes which followed them in to drop their loads on the fires. The British countered with false fires set up outside of London in the path of the advancing bomber. These fires were piles of waste which were prepared in advance. The false fires were called Starfish. After the Germans dropped their incendiaries, several Starfish would be ignited. The bombers, seeing the Starfish first, would drop their bomb loads on them, thinking them the target. When the British first used Starfish, as many as 95 percent of the bombs were dropped on these false targets. When the Germans realized the ruse, they dropped their bombs on the second area of fires. Again Starfish was used, but on the other side of the city, so that the second fire was not the target. The results were a division of droppings of the Germans' bomb loads. About 30 percent of the bombs we. been usually dropped on Starfish.

Several important ... applicable to ECM were learned in this navigational phase of the war. In the first place, the Germans carried on their experiments in range of the British monitors, or tried out new systems in advance on a small scale. As a result, Britain had a big advantage in time for developing countermeasures. The Germans also set up their beams early in the afternoons, giving the British opportunity to fly down the beams and determine the targets to be bombed that night.

Libyan Campaign. British jamming of the radio channels began in the Libyan campaign, starting 18 November 1941. The British had not used jamming prior to this time because of their fear of retaliation by Germany. The communications channel jammed was the German tank radio on 27 to 35.5 Mc. To carry out the jamming, Wellington bombers were equipped with 50-watt transmitters covering the frequency range. The frequency of each transmitted signal was modulated to prevent giving a steady tone. A special antenna was made which could be towed by the bomber and which was about 3 inches in diameter and about 9 feet long. The jamming operation was very successful. Stations more than 2 miles apart were unable to communicate with each other. Those closer together than 2 miles had very erratic operation. The jamming equipment consisted of an oscillator fed to a power amplifier. The frequency was varied by mechanically rotating one element of a condenser in the oscillator circuit. The British neglected to provide fighter protection for their jamming planes, and as a result, the jamming was soon interrupted.

British Jamming of German SSV Equipment. The next operational jamming was done against the German SSV (Shore-to-Surface Vessel) located off the French coast. Each time the British brought a convoy through the English Channel, German radar-controlled guns would fire on the convoy. The British set up land-based jammers across the channel and jammed this

radar effectively. Some important facts relative to ECM came out at this time. The Germans had all their equipment on approximately the same frequency, thus making it easy to jam. The first time the British jammed, the Germans turned off their radar, thereby divulging that jamming was effective. Later, the Germans staggered the frequencies used, making jamming more difficult. The Germans started jamming in February 1942 and eventually built a jammer for every English radar unit, including gun laying.

2.1.2 Initial United States Efforts in Electronic Warfare During World War II

In an effort to supplement the already overburdened Army and Navy research programs, OSRD, the Office of Scientific Research and Development, was established by executive order in June 1941. It contained 3 major branches, of which one, NDRC, was in turn divided into 19 divisions. Administrative functions and liaison with the Army and Navy were performed by the headquarters of OSRD.

OSRD was founded for the purpose of carrying on research in support of the Army and Navy on an emergency basis. It was provided with its own funds for financing research contracts with universities and industrial organizations throughout the country to fulfill the urgent wartime requirements of the Army and Navy.

2.1.2.1 National Defense Research Committee (NDRC)

NDRC was the branch of OSRD which was concerned principally with the physical sciences, and it was within the structure of NDRC that the OSRD countermeasures program was carried out. Each of the 19 "divisions" of NDRC was headed by a committee. The Division 15 Committee, under Dr. C. G. Suits, administered the countermeasures program. Division 15 contained several major subdivisions, including an office at Schenectady, New York, responsible for tube contracts with various manufacturers; an office in New York City, which administered contracts for countermeasures; and an office in Cambridge, Massachusetts, responsible for the administration of contracts in the Boston area, including the contract under which RRL, the Radio Research Laboratory at Harvard University, was operated.

2.1.2.2 Coordination of the Program

The Army and Navy coordinated ECM program was carried out by NDRC in cooperation with the various laboratories and bureaus of the Armed Services. Chief among the agencies representing the Services were the Aircraft Radio Laboratory at Wright Field, Ohio, the Naval Research Laboratory, the Office of the Chief of Naval Operations, the Bureau of Ships,

the Bureau of Aeronautics, the Office of the Chief Signal Officer, and the Signal Corps Engineering Laboratories at Fort Monmouth, New Jersey. Each of the service laboratories involved had its independent program of research and development, but coordination among laboratories was maintained. For example, one of the Services might express an operational requirement for a particular type of equipment. The initial investigation might be carried on solely by NDRC or by one of its laboratories in collaboration with the operating Service. Later, development and procurement would be handled by the Service involved. In general, the operational need for particular development was established and expressed by the Service on an informal basis, usually in meetings attended by the personnel of the various laboratories and agencies concerned.

2.1.2.3 Radio Research Laboratory

Work in the field of radar countermeasures was started prior to the establishment of the Division 15 organization when, in 1942, a small group was set up in Division 14's Radiation Laboratory at M.I.T. under the direction of Dr. F. E. Terman, for the purpose of developing jammers to use against enemy radar and also of developing antijamming devices for incorporation in our own radar.

The general problem involved was the development of means whereby the effectiveness of the enemy's radar equipment might be nullified. It became evident in the early days of the war that radar was not only a very useful weapon but that it was useful both for ourselves and for the enemy. It was also evident that it was a very vulnerable weapon. On the one hand it appeared prudent to take steps to make this weapon as useless to the enemy as possible in case he should attempt to use it against us, and on the other hand it seemed practically essential to do something about the vulnerability of our own radar in case the enemy should attempt to jam us.

It soon became clear that the radar countermeasures program was much too extensive to be carried on as a part of radar development activities. Steps were accordingly taken soon after the establishment of the radar countermeasures group to move it to Harvard, where it was established as the Radio Research Laboratory, operated exclusively for Division 15 of NDRC by Harvard University under the direction of Dr. Terman. During its approximately 3½ years of existence, Radio Research Laboratory grew to a peak strength, in August 1944, with some 810 persons.

2.1.3 Initial Tactical Applications

Following the issuance, in July 1941, of a Presidential order to the U. S. Navy to attack all enemy submarines, the Navy established a complex of

shore DF stations with the technical assistance of the NRL. The successful "Wolf Pack" tactics developed by the Germans for the exploitation of their attacks on convoys required that high-frequency communications be employed in making a rendezvous for the pack. The DAJ hf direction-finding equipment was guided into production by NRL engineers and formed the backbone of the Navy's shore DF program. This equipment was produced by the Bureau of Ships in large quantities, based on improved NRL designs. A shipboard counterpart of the shore-based DF was also produced. Improvements in the model DAQ equipment were worked out jointly with the British, and a program was also initiated to train personnel to install and operate this new equipment for the Navy. Unquestionably, these equipments, developed jointly by the Admiralty and NRL, and produced separately, spelled the doom of the "Wolf Pack" tactics of the German submarines, even though their transmissions grew shorter and shorter as time went on. The initial location of the enemy submarine was the most important phase of the antisubmarine warfare success achieved by the Navy. Sonar and radar were brought into play for the kill, but the oceans are large and the small number of antisubmarine warfare units urgently needed clues on where to hunt. These clues were always provided by the hf DF operators.

The use of jamming during the escape of the German warships *Scharnhorst* and *Gneisenau* was perhaps the first serious indication of the importance of countermeasures in the naval warfare picture. General Martini, the Luftwaffe's Director of Communications, had completed a series of jamming exercises which, when directed at the British radars at dawn each day, had resembled atmospheric interference. The length of jamming was increased each day until by 11 February 1942 the British radar operators were thoroughly accustomed to this particular type of interference, which they reported normally as caused by atmospheric conditions. This cleverly executed plan covered the escape of the two German warships up the Channel from Brest.

Later, German production jammers were employed extensively to ring the Mediterranean. Allied shipborne metric radars had their scopes completely jammed from the time they entered the Mediterranean until they left. German jammers passed the Allied ships from one CM group to another, keeping the vessels constantly under the devastating effects of the metric jammers.

The Germans also demonstrated their ability to take the offensive in ECM when they intercepted and took control of a group of U.S. radio-controlled boats also in the Mediterranean. On this occasion the boats were sent in tight circles; thus expending their fuel harmlessly. It was a disheartening experience for the Navy to encounter such proficient employment of electronic warfare by the Germans.

2.1.4 The Radar Countermeasures Problem

When bombing raids over Germany were started, the extent of the German radar development was immediately realized. Not only were there enemy interceptor airplanes equipped with airborne radar capable of locating our bombers through the heaviest overcast, or at night, at distances up to 10 miles, but also the damaging antiaircraft fire into which our airplanes ran was found to be directed by a very effective F-band gun-laying radar known as the *Small Wursburg*. A similar (500 Mc) radar, the *Giant Wursburg* (see Figure 2-1) was used by the Germans for fighter control, and they depended

Fig. 2-1. German Giant Wursburg.

on a 125-Mc set, the *Freya*, for early warning. In the initial days of our bombing efforts against Germany, losses we   extremely high due to these enemy weapons.

The first report of German use of radar was in 1941. Electronic reconnaissance flights indicated that the Germans were using radar to guide their anti-aircraft defenses. Eventually, photographic coverage was focused on an installation on the French coast at the town of Bruneval on the English

Channel coast opposite England. This radar watched the coastal shipping in the Channel. The Germans also coupled the Giant Wurzburg radar with some long-range coastal batteries which were effectively scoring many hits on the channel shipping. This was one of the first more modern radars since it combined fire control with its search function. Its antenna was a 25-foot-diameter parabolic reflector which operated at a frequency of 570 Mc. The British were indeed worried about this radar and the rapid appearance of the other similar radars. In February 1942 they decided to move in against the radar at Bruneval. A very carefully planned commando raid was organized by the British. The risk was great and the possibility of success was highly questionable. The British succeeded in capturing the receiver, parts of the indicator, and other vital portions of this Wurzburg. These units of the radar were set up in England and carefully studied. As a result, the British were able to examine critically their jamming effectiveness by noting the performance of their equipment against this practical mock-up of the German Wurzburg radar.

The problems involved in developing effective countermeasures which could be applied against this enemy radar equipment were not as simple as they might seem on the surface. In the first place, there was no assurance that enemy radar development would stop with the existing equipment. On the contrary, it was expected that, as in this country, Germany would proceed with radar development activities in an effort to improve the definition and performance of their equipment. In general, such improvements could come about only through an expansion of the frequency range. It was therefore probable that it would be necessary to develop jamming equipment for use against enemy radar equipment operating at much higher frequencies than those which had been detected already. The only safe way to cope with the situation was to develop jamming equipment capable of being tuned to any conceivable frequency which the enemy might employ.

Second, although vacuum tubes, which had been developed for use in radar sets, were capable of providing large power outputs at high frequencies in pulses lasting a few microseconds, radar jamming techniques required a different approach. A radar jamming signal must be emitted continuously; if it is not, the "pip" which reveals to the enemy the location, on his radar scope, of our planes will show through in the intervals between jamming signals. The development of new tubes was therefore required.

Third, it was not enough merely to develop jamming transmitters. In order to operate them successfully, it was necessary to know the location and the frequency of the enemy radar set, so that the jamming system might be

properly tuned, turned on at the proper time, and in some cases, pointed in the right direction.

That the development of a successful radar countermeasures program would involve extensive effort along a number of lines was, therefore, quite evident. As finally established, it was planned that a complete line of jamming transmitters, receivers, and direction finders would be developed to cover the entire known or possible radar spectrum. The size of the job involved in implementing this program may be judged by the fact that, when it was initiated, even the fundamental techniques of generating radio-frequency energy at the frequencies involved were imperfectly known. Following the development of the new types of vacuum tubes required for this purpose, it was necessary to build not one, but dozens, of types of equipment, each one capable of covering a small part of the radio spectrum.

2.1.4.1 Vacuum Tubes

The tube problem was solved in a number of ingenious ways. It was discovered that the magnetron, invented by Dr. A. W. Hull in 1921, was an efficient generator of high-frequency energy in the microwave region. At once, activity started which resulted in the development of a great diversity of magnetrons suitable initially for use in radar sets, but later adapted to radar countermeasures. Division 15 placed contracts with various contractors for the development of a complete line of CW tubes capable of covering the radar spectrum at various power levels. Later, these tubes were incorporated in 100-watt jamming transmitters.

The tube program included the development of a line of split anode CW magnetrons designed for a nominal power output of 150 watts in the frequency range from 90 to 1200 Mc. Successful experimental models of these tubes were completed, and they formed the basis of certain jamming transmitters which were developed at RRL, such as the AN/APT-4 and TDY and TDY-1 equipments, later placed in procurement. Typical split-anode CW magnetrons received the type numbers ZP590, ZP579, and ZP584 (General Electric Company). Another tube program contemplated the development of a line of 1-kw CW magnetrons for frequency ranges above 1000 Mc. A few samples of these tubes were produced, but the majority never passed beyond the development stage in the tube laboratories.

Another program for developing low-power S-band CW magnetrons was undertaken. One of these tubes, the QK-44 (Raytheon) was used in the S-band RRL F-5100 jamming transmitter, which was completed just prior to the end of the war and was the forerunner of the AN/ALT-8 and AN/APT-16 jamming equipments later produced.

Tubes of the so-called "conventional" type, but adapted to operate in the

microwave region by reducing the spacings between the elements to almost microscopic dimensions, were developed by several tube companies. These tubes were called "lighthouse" or "oilcan" tubes on account of the similarity of their appearance to the objects named. They turned out to be very useful in both transmitting and receiving applications where lower power was required.

Another tube extensively used in receivers on account of its wide tuning range was the reflex klystron oscillator. With the use of tubes of this type, a line of search receivers was developed which, at the end of the war, was capable of covering practically the entire radar spectrum. And finally, the "resnatron," which was capable of CW power outputs of the order of 75 kw at a frequency of 500 Mc was developed and later built into a mobile jamming transmitter ("Tuba") in seven Army trucks, which was used for a few weeks near the end of the war to clear a path for British bombers flying across the English Channel.

2.1.4.2 Antennas

The antenna problem was difficult. Transmitters and receivers alone were useless without antennas. And, at the frequencies employed, conventional antennas were not satisfactory. An antenna research program was undertaken, as a result of which the conventional "wire" and "whip" antennas of radio broadcasting speedily shrank down into "stubs" 6 to 30 inches long, adapted to radiate the extremely high frequency energy with maximum efficiency. At still higher frequencies the stubs developed vertical slits along their sides and became "split cans" and "slots." In the microwave region it was possible to dispense with antennas entirely in the conventional sense and spray the energy out of a horn.

As part of the antenna development program, extensive antenna-pattern measurements were made by The Ohio State University Research Foundation. This work also included measurements of the echoing patterns of model aircraft. The problem of vhf and uhf antennas was studied with particular emphasis on means for reducing pattern distortion due to the characteristics of the aircraft. The result of this research was the "Stingaree," a relatively broadband doublet antenna which was designed to be trailed behind the aircraft using a coaxial fitted cable as a tow line, and which was thus removed from the immediate vicinity of the aircraft.

2.1.4.3 Window

One of the most successful jamming expedients was "window" or "chaff," the code name applied to the use of tuned aluminum foil strips which, thrown out in large bundles from our bombers, created a "radar smoke

screen" in the scopes of enemy radars and completely masked the bombers from radar detection by the enemy. Chaff was used quite early in the war by the British, who employed flat, covered metal cards printed with propaganda messages to mask the real purpose of the material. In its original form, chaff was cumbersome and heavy; consequently, the principal objective of the wartime chaff research groups was to refine this extremely effective "confusion" technique by reducing the dimensions of the strips and tuning them to the enemy radar frequencies. All of the early chaff was actually cut with scissors by squadron personnel, from sheets of foil in the field. Large quantities of the improved chaff were used, both in the European theatre and in the Pacific, with great success in combination with electronic
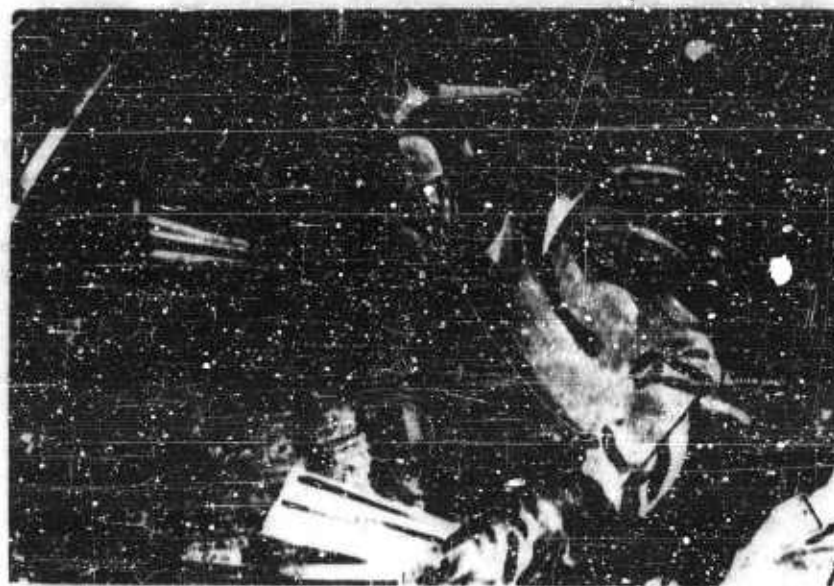

Fig. 2-2.   Dispensing chaff.

jamming (see Figure 2-2). Ultimately, some 20,000 tons of chaff were produced.

The Luftwaffe in desperation threw roughly 4000 engineers into the breach to solve the antijamming and window problem that was plaguing its radars. This effort represented roughly 90% of their uhf engineers. In the United States only about one-tenth this number were engaged in developing countermeasures against radar. Finally, the Luftwaffe announced a public competition with prizes totaling 700,000 Reichmarks (free of all taxes) for the best

solution to the problem of window. In their rush to save the Wurzburgs, the Germans were distracted from the development of microwave radar which was already being exploited by the Allies.

Tests of the completed chaff material under operational conditions soon disclosed that means would be required for its automatic and homogeneous ejection. Dispensers were ultimately developed which, in conjunction with pure foil aluminum strips, were able to eject large quantities of chaff in a short period of time without "bird nesting"—the gathering of the material into clumps.

The low-frequency radars used by the Japanese required a window technique different from that employed in the European theatre. In the 100-200-Mc range, chaff bundles became long and bulky and difficult to handle. This led to the development of a type of radar reflector known as "rope," which consisted of 400-foot lengths of thin aluminum tape $\frac{1}{2}$ inch wide.

In the high-frequency range, "corner reflectors," known as "angels," were worked on but found ineffective.

2.1.4.4 World War II Direction Finders

A number of types of direction finders were developed as a part of the World War II ECM program, then known as the "RCM" program. One, a homing system for use in fighter planes was known as the AN/APA-48 equipment. This equipment, which was being placed in service when the war ended, was to have been used to locate Japanese radar-equipped "Snooper" planes, which made a practice of tracking our task forces from a distance just outside the range of U. S. ship radars.

A simple direction-finding system, known as the AN/APA-24, using dipole antenna elements, was developed for the lower frequency ranges. This system covered a frequency range from 70 to 400 Mc, was provided with a series of plug-in interchangeable heads, and was designed for mounting on aircraft such as the PBM. The APA-24 was a null-type device, which required manual adjustment to determine the direction of a received signal.

A more complete system, in a different frequency range, capable of simultaneously locating a number of signals without manual adjustment, was the AN/APA-17 rotating-reflector direction finder. By the end of the war, heads for this equipment had been developed which permitted it to operate over the frequency range from 200 to 10,000 Mc.

The shipborne counterpart of the AN/APA-17 was the DBM direction finder. This equipment covered a frequency range from 150 to 5,000 Mc, using two heads, and extension of the range to 10,000 Mc for airborne applications was undertaken. Subsequently, a 4000-10,000-Mc head was developed by NRL.

The largest part of the ECM program in the Signal Corps was centered in the Radio Direction Finding Branch at Eatontown Signal Laboratory (ESL), Fort Monmouth, New Jersey. Here, direction finders such as the AN/TRD-5, a V-T fuze jammer locater, covering 60-300 Mc, and the AN/TRD-4, AN/TRD-10, and AN/TRC-2 communications DF systems were developed. The AN/PRD-1 covering 500 kc to 30 Mc was the most advanced portable loop direction finder. The AN/TRD-2 was the first instantaneous visual DF. This covered the frequency range 1.5-18 Mc and used the BC-342 receiver.

2.1.4.5 Test Equipment

In the interest of assuring that equipment delivered to the field would be properly used, the development program included a number of projects for the development of test equipment. Among equipments which were completed in the laboratory and turned over to manufacturers for production were the BC-1255A heterodyne-type frequency meter for the low-frequency jamming transmitters Dina and Mandrel; the TS-92/AP alignment indicator for Dina, the TS-47/APR test oscillator for the AN/APR-1 and AN/APR-4 receivers; and the TS-131/APT transmitter output indicator. Because of production delays, many of these equipments were not actually delivered until late in the war.

2.1.4.6 Antijamming

Representative types of U. S. radar were set up at RRL, NRL, and ESL for making studies of vulnerability of particular types of radar. This program resulted in the production of a number of training aids, reports and technical manuals, motion pictures, training signal generators, and training jammers, which were used to carry out demonstrations of countermeasures at Army and Navy operating sites and training schools. Several antijamming devices for incorporation in standard radar equipment were also developed and made available for use.

2.1.4.7 World War II Transmitters and Systems

In the last analysis, all of the work of the World War II countermeasures program was aimed at making it possible to use active or passive jamming against the enemy. The passive jamming phase of this objective has already been considered briefly in Section 2.1.4.3.

In considering the active jamming question, many difficulties arose. There was, primarily, the question of tuning the jammer to the frequency of the device being jammed. This required the use of search receivers, preferably of the rapid-scan or wide-open type. Tuning of the trans-

mitter was then necessary. The design of the transmitter itself was an important factor. Noise modulation was, in general, employed, but it was possible to employ either wideband barrage jamming or narrowband spot jamming. Barrage jamming was generally difficult to achieve, because of the relatively narrow bandwidth of available noise sources and the fact that it was necessary to modulate the rf generator with the noise source. On the other hand, barrage jamming possessed the important advantage that exact tuning of the transmitter was not required. With sufficiently broadband barrage jammers it was possible to locate a multiplicity of jamming transmitters in the aircraft of a flight (on the ground or on shipboard), to tune these transmitters to adjacent sections of the spectrum, and to eliminate tuning en route entirely.

The earliest transmitter developed by the Radio Research Laboratory was the Carpet I-Ia, AN/APT-2, AN/SPT-2, designed for the frequency range from 400 to 720 Mc, and adapted for use against the German Würzburg gunlaying radar sets. This transmitter was designed for a power output from 5 to 10 watts.

The Carpet jammer, Carpet III, AN/APQ-9, AN/SPT-5, AN/UPT-1 (RRL F-2500) for the power range from 15 watts to 80 watts supplemented the Carpet I-Ia equipments.

The Dina (direct noise amplifier) transmitter, AN/APT-1, AN/SPT-1, for the frequency range 90 to 220 Mc, having a power level of 10 to 30 watts, was used against the *Freya* early warning radar (125 Mc).

The Rug transmitter, AN/APQ-2, AN/SPT-4 (RRL F-1500), covered a higher frequency range (200-500 Mc) at the 20-watt level. Carpet IV, AN/APT-5 (RRL F-5500), again at 20 watts, covered the range from 350 to 1200 Mc.

The lighthouse tube transmitter, AN/APT-9 (RRL F-4800), had a range from 300 to 2500 Mc, again at 20 watts. The RRL F-4500 jammer covered a frequency range from approximately 2700 to 4200 Mc at 20 watts. This last experimental jammer was, however, never completed but was replaced by the S-band jammer F-5100.

Quite early in the war, the research and development work at the 20-watt power level had been completed, to the extent that the spectrum from 25 Mc to 4200 Mc was pretty well covered by jamming equipment of one kind or another, including Carpet I-Ia, Dina, HF Dina, AN/APT-1, Mandrel, AN/APT-3, Rug, Carpet IV, the F-4800 jammer, later the AN/APT-9, and the F-4500 jammer.

Coverage at the 150-watt level was also fairly complete since it extended from 90 to 1200 Mc, including the Dina amplifiers (RRL B-4100, B-2800), the split anode magnetron transmitter AN/APT-4 (RRL F-3400, F-3410),

Fig. 2-3.   Magnetron jamming transmitter AN/APT-4.

and the magnetron transmitter F-4400. The Navy also developed the TDY, TDY-1 split-anode magnetron 150-watt transmitter, utilizing the same tubes employed in the AN/APT-4, the ZP579, and ZP590. Figure 2-3 shows the AN/APT-4 transmitter.

On the other hand, transmitters for the 1-kw level were limited to relatively narrow segments of the spectrum, and the 10-kw and 40-kw levels were represented by research projects involving specific types of tubes, the ZP595 magnetron and the Model 21 resnatron. This tapering off of frequency ranges with increasing power was, of course, to be expected in view of the technical limitations existing at the time.

The first operational use of production "Carpet" jammers was on October 8, 1943, only about eighteen months after the start of the program. Carried on this date in 68 aircraft of two groups of the 8th Air Force during a raid on Bremen, the "Carpets" resulted in a 50 percent reduction in losses for the protected groups.

In addition to the transmitter projects discussed above, and other such projects not mentioned, three systems projects were established. These were the AN/APQ-20 and AN/APQ-27 systems, which used the F-5100, 50-watt S-band jamming transmitter, in anticipation of enemy developments; and the XMBT "Elephant" 1-kw S-band jamming system, S-9000, designed for ship-

borne use. After the war, the F-5100 project ultimately resulted in the AN/APT-16 jamming transmitter, and the "Elephant" led to the high-power jammers AN/MLQ-2, AN/MLQ-7, and AN/SLT-2 and -3.

A major naval effort at the Naval Research Laboratory was concerned with the employment of countermeasures against the Germans' HS-293 glide bomb. Early in the war the use of this bomb offered severe resistance to our naval units in the Mediterranean. This weapon was first used operationally to sink the Italian battleship *Roma* as she was attempting to escape to join the Allies. It was used also to sink the British ship *Warsprite*, and it damaged the U. S. Cruiser *Savannah*. It was a report by the crew of the *Savannah* that gave the Naval Research Laboratory the clue that indicated that the Navy was faced with a guided weapon. Later one of these HS-293 glide bombs sank in a near miss in shallow water off Libya and was recovered by the British. Meanwhile NRL engineers were called to work around the clock on the research and development of equipment to counter the threat of this bomb. Experimental equipment, fitted on ships in only six weeks time, was used to intercept, record, and analyze the guided-bomb control signals. It was desired to obtain shipboard recordings, while the ship was under attack, of the radio control signals for complete laboratory analysis. In practice, two of the four control tones employed were above the audible frequency range, and also were above the frequency range of World War II recorders. Only after a series of ingenious maneuvers were these radio frequencies and their associated tones successfully located and accurately analyzed. Soon after the first control signals were analyzed, two destroyer escorts, the U.S.S. *Davis* and the U.S.S. *Jones*, were supplied with experimental NRL equipment which so successfully jammed the guided bombs that the effectiveness of the weapon was very nearly neutralized. After this jamming program gained full momentum, no major fleet unit was sunk by the glide bombs.

Following these initial successes by the Navy with the Naval Research Laboratory experimental equipment, Airborne Instruments Laboratories at Mineola, New York, undertook a project for the development of production jammers for use against the German HS-293 glide bomb for the Navy. This equipment, known as the MAS jamming system, involved the development of several types of multiple-channel receivers and manually tuning spot jammers, and some automatic search receiver-spot-jammer combinations. A crash program was also undertaken at RRL at the time when these German glide bombs were first used in the Mediterranean, in March 1944. RRL was asked to produce on a crash basis a quantity of 30 jamming systems for use against the glide bombs. In response to this request, RRL converted a quantity of AN/ARQ-8 transmitter-receivers, which were then in crash production for the Air Force, and made them suitable for the Mediterranean operation.

Developments in the theater of operation, however, changed the operational needs, and none of these equipments saw service in the field.

Airborne Instruments Laboratories also worked on an airborne jammer (AN/ARQ-11) having a power output of about 1 kw. Ten of these units were delivered to the Air Force as prototypes of a high-power airborne jammer. In the V-T fuse jamming field, the Signal Corps Laboratory, Fort Monmouth, New Jersey, developed the AN/TRT-2, a convoy jammer with approximately 100 watts output, swept CW, covering the frequency range 75-200 Mc.

The Westinghouse Research Laboratories was responsible for the development of a 10-kw ground-based communication jammer known as "Ground Cigar," for the frequency range from 38 to 42 Mc. Other work in the field of communications countermeasures and communications antijamming was done by the Bell Telephone Laboratories, by the Federal Telephone & Radio Corporation, and by the Radio Corporation of America.

2.1.4.5 World War II Receivers

The fundamental success of any countermeasures operation must be dependent on the type of receivers available, because, before recourse can be had to active jamming techniques, it is necessary to determine the location and frequency of the electronic systems to be jammed. The primary emphasis of the countermeasures program, when it was set up initially was, therefore, on the development of receivers.

The Navy, faced with locating German submarines equipped with radar, met this threat with simple tunable detector video receivers developed at the Naval Research Laboratory. These first receivers had stub antennas that extended through the skin of the aircraft, and the base end of the antenna inside served as a tuning cavity to determine the frequency. Tuning was effected by sliding a shorting bar up and down the cavity until a signal was detected. The later quantity-produced sets were designated ARD-1 and ARD-2 and included circuitry for determining prf and pulse length.

These same sets were also used by the Naval Patrol aircraft in the Pacific to locate the Japanese radars located throughout the islands. They were also used to locate the first Japanese shipborne radars used in the Pacific fleets. These simple receivers filled an important operational gap for the Navy prior to the production of replacement receivers.

Subsequently, the receiver program comprised, principally, work on three receivers, the AN/APR-1, the AN/APR-4, and the AN/APR-5A. The AN/APR-1 receiver and the AN/APR-4 receiver were, in general, similar, except that the AN/APR-4 receiver permitted a choice of i-f bandwidth. Most of the work on these two receivers involved the development of three

different types of tuning units, identified by the suffixes A, C, or D, following the project number. The A units were, in general, motor drive units for the AN/APR-1 receiver; the C units were manual tuning units for use with the AN/APR-1 receiver; and the D units were designed for use with the AN/APR-4 receiver, and were equipped both with motor drive and sector sweep. The AN/APR-5A receiver was designed to cover the higher frequency range from 1000 to 6000 Mc. These receivers went into large quantity production and were extensively used in the field.

In addition to the work on the AN/APR-1, the AN/APR-4 and AN/APR-5A receivers, the program included another receiver, the AN/APR-2, known as the "Autosearch" receiver. Work was also done on "wide-open" receivers, such as the "Spud," the "Zero Catcher," and the "Boozer," but these wide-open receivers never saw extensive use due to changes in the operational requirements

The frequency coverage afforded by RRL-developed receivers which reached production was from 40 Mc to 6000 Mc. At the end of the war, work was just starting on the AN/APR-9 receiver, which was intended to give greater frequency coverage and enhanced performance. Work on the AN/APR-9, was transferred to the Airborne Instruments Laboratories at Mineola where the development was completed under Navy sponsorship.

In the V-T fuze field, the Signal Corps Engineering Laboratory, under contract with Airborne Instrument Laboratories, developed the AN/ARR-24, a lightweight, rapid-scan superheterodyne covering the frequency range 60-300 Mc in one band. This was completed in 1944 but was not used since the Germans did not possess operational V-T fuzes. At the end of the war, work was started at SCEL on the AN/TLR-9 and -10 V-T fuze receivers covering 60-1050 Mc with high sensitivity and probability of intercept.

2.1.4.9 Evaluation of ECM

Before the end of the German war, following construction of prototype models, most of the equipment developed by the joint Army-Navy-NDRC program had been placed in procurement, and nearly every bomber in the 8th Air Force had been equipped with at least one, and in some cases up to four, Carpet jamming transmitters. Results were anxiously awaited. Representatives of Radio Research Laboratory stationed at the Division 15 British Laboratory, ABL-15, at Malvern, England, and at other strategic points in the operational theatres, made operational analyses to determine the effect on the enemy. Soon after the equipment came into full use against the enemy, losses in the 8th Air Force began to drop. While initially it had been expected that an appreciable fraction of our bombers would fail to return from raids over Germany, the percentage of losses was now reduced to a very low figure.

These operational analyses were not conclusive. The Germans had lost much of their antiaircraft equipment in bombing raids, and the Luftwaffe had practically stopped operating. The weather had also changed during the period when statistics were accumulated. It was known that the losses had decreased, but it could not be proved that the countermeasures program had achieved this result, or had even helped.

It was not until after VE Day that the truth was learned, from the lips of the Germans themselves. Within a week after hostilities had ceased, almost the entire staff of the American-British Laboratory of Division 15, in England, had scattered to various points within Germany. Soon reports began to come back. The records of the Reichforschungsrat (the German OSRD) had been examined, and it had been determined that the American-British program was a spectacular success. German scientists had been interviewed, who confessed themselves baffled by our countermeasures activity. Representatives had talked to radar officers in the German ground forces, who declared that our countermeasures had rendered German antiaircraft radars useless.

The sum total of the investigations in Germany confirmed the view that the ECM program had been a success. It was true, to a very considerable extent, that the countermeasures gear had been a major factor in the reduction in losses. The entire Nazi radar network, according to the people operating it, had been reduced to about one-fifth of its normal effectiveness. Fairly early in the war the Germans had learned to depend almost entirely on radar for antiaircraft gun control, because it gave a much more accurate range and was reliable in all kinds of weather. When the 8th Air Force began using window and electronic jamming, the German antiaircraft crews had been blinded. Try as they might, they had been unable to determine the location of our flights through the dazzling glare of their radar scopes. Orders had been issued to continue firing in spite of the interference, in order not to reveal to us the fact that our countermeasures had been successful. Then, so poor was the record of planes shot down under these conditions, that these orders were replaced with orders not to fire at all unless good visual aim could be obtained—orders equivalent to abandoning radar antiaircraft control entirely.

In the German laboratories, scientists had been at work attempting to lessen the vulnerability of their radar equipment ever since the British had dropped the first window in the raid over Hamburg in 1943. After the German bombing raids on England in 1940, Hitler had thought the war was won and had ordered the demobilization of a great part of the German scientific effort and the induction of the scientists into the army. With the Hamburg raid, and the capture of one of our advanced airborne radar sets, the Ger-

mans had seen the error of this decision and had immediately reconstituted the scientific organization. At the end of the war, the laboratories had been operating at full capacity, and about half the German scientific effort in the field of electronics had been directed against our countermeasures activity. So large a force, in fact, had been engaged in this work, that efforts along other lines of scientific war developments were neglected and it was the opinion of investigators that the countermeasures program had not only nullified the German antiaircraft fire but the entire scientific program in general, through the preoccupation of the German scientific organization with the countermeasures program.

Prior to D-Day, the U. S. Naval Research Laboratory had a team in England working with the Admiralty scientists, preparing especially fitted ships for the countermeasures role in the invasion. A total of some sixty ships were fitted. NRL engineers in England fitted ten ships and an additional five were fitted in the United States by NRL personnel. These ships were configured with the countermeasures for the glide bomb. In addition, NRL engineers worked with the British in developing the plans for countermeasures deception. Small boats were fitted with a series of towed deceptive "targets" which were made of chicken wire in the form of a corner reflector. Other boats towed balloonborne deceptive reflectors. These units were towed close ashore on the French coast on the morning of the invasion and were responsible for the early report by the Germans that the attack was coming near the Pas de Calais and Boulogne areas many miles away from where the landings were actually made. In spite of their dangerous deceptive role the small boats returned to England without loss in the face of heavy radar-directed shore-based gunfire. The ships equipped with glide bomb countermeasures did a commendable job since they protected the fleets from the hundreds of glide bombs loosed during the attacks on the invasion fleet.

Window was also used extensively during the Normandy invasion. Army Air Force B-24 aircraft with window flew a racetrack pattern over the English channel near the coast of Normandy before and after the invasion. B-24 aircraft equipped with "Carpet" jamming transmitters participated in the diversionary action against the Pas de Calais area.

2.2 Electronic Countermeasures Since World War II

When the Japanese war ended, in August 1945, steps were immediately taken to place in effect plans which had already been worked out for the demobilization of the Radio Research Laboratory. Certain projects for which there was no further need were terminated immediately. Other activities which were near completion and which had continuing values, were completed in order to preserve those values. Certain other projects which

evidently could not be completed in the near future, but which had considerable long-range importance, were transferred to laboratories of the Armed Services, such as the Naval Research Laboratory and the Aircraft Radiation Laboratory. All laboratory work at the Radio Research Laboratory was stopped by November 1, 1945. By January 1946 only about 250 persons were left on the RRL payroll, and of these, less than 10 percent were scientific personnel, all of whom were engaged in working on reports.

A limited amount of research and development effort took place at the end of World War II and continued for a year or so thereafter. Many of the personnel in the Armed Services who had been engaged in this effort returned to civilian life, and the greater part of the equipment produced for ECM purposes during World War II was sold on the surplus market.

In the late 1940's, ferret activities showed that a radar net was being built rapidly around the Iron Curtain. The ECM program was thereupon re-established, starting, on the part of the Navy and Air Force, with a small research and development effort, schools, and continued ferret activities. World War II equipment was obtained from warehouses and the surplus market, and the need for newer and better equipment was recognized.

From 1947 to the present date, an extensive countermeasure research and development program has been carried on under the direct sponsorship of the Armed Services. Work has been done, in general, by Service Laboratories or Universities and industrial organizations under prime contracts with various branches of the Services.

Because the program is now so widely decentralized among the various branches of the Services and among a multitude of contractors, it is difficult to provide a full and complete picture of the current situation, particularly insofar as the availability of production equipment is concerned. No attempt will be made to present a fully comprehensive and exact analysis of the situation; but rather, especially significant developments will be reviewed without tabulation, and not necessarily in chronological order.

Performance requirements for ECM equipment have now become much more exacting. While, during World War II it was possible, under favorable conditions, for an ECM operator to ascertain the location and frequency of a signal to be jammed and to tune a jamming transmitter to this frequency manually, such procedures are now out of the question. Aircraft speeds have increased manyfold, reducing the time available for manual operation. The density of radar signals may be expected to be much greater in any future conflict. Weight limitations make it extremely desirable to dispense with operators entirely in aircraft, and on the ground and on shipboard the necessity for rapid action also favors the abandonment of manual control. The postwar trend has accordingly been in the direction of the development of

more sophisticated jamming techniques. World War II jamming equipment was, in general, of the "brute force" variety, lacking in precision and finesse. We now have in production or development automatic jamming systems of the automatic search and lock-on type, range gate and doppler gate pulloff repeaters, inverse gain repeaters, and other equipments utilizing novel techniques to achieve automatic or semiautomatic operation. The B-58, B-52, and B-70 all include an electronic warfare defensive position in the basic design.

It is noted that, on account of the complexity of such systems, it is, in some cases, difficult to meet weight and size requirements. It appears, therefore, that there may still be a very useful application for conventional wideband, untuned, noise-modulated, barrage-jamming transmitters. This problem is aggravated by the fact that the power requirements of jamming equipment have continuously increased in step with the increased output of the radar or other systems to be jammed.

The research and development program is following, in general, the same lines which were established during the war. In other words, the program is still directed toward the development of equipment for radar countermeasures, communication countermeasures, guided-missiles countermeasures, and proximity-fuse countermeasures. Important changes in emphasis have, however, taken place since the war. While a major amount of effort is still being expended on radar countermeasures, the portion of the program devoted to communication countermeasures and proximity-fuse countermeasures has increased materially. Additional emphasis is also now being placed on guided missiles countermeasures, principally through the establishment of the Electronic Defense Laboratory of Sylvania Electric Products, Inc., at Mountain View, California, under the sponsorship of the Signal Corps. Considerable effort is taking place in the development of techniques and equipment to support the ELINT (Electronic Intelligence) programs of the various services.

From the standpoint of structure, the program is also divided into about the same categories which existed during World War II. We still have work on transmitters, receivers, antennas, direction-finding equipment, testing and training equipment, window, and antijamming devices. Here, however, even greater changes in emphasis have occurred than in the case of the various applications toward which this program is directed. Due to the fact that operational requirements are now considerably more stringent, and due to the necessity for integrating the jamming equipment used by the operating forces with other types of electronic gear available for their use, an increased amount of effort is now being expended on the development of complete systems adapted to do the jamming job more rapidly and effectively.

With the objective of satisfying these new operational needs, requirements have been altered to call for more rapid acquisition times, higher jamming power levels, enhanced sensitivity in receivers, the use of directional antenna to obtain increased radiated power, improved receiver resolution, and expanded frequency ranges.

2.2.1 Postwar Transmitters and Systems

Shortly before the end of the war, work on the first S-band noise-modulated jamming transmitter was completed. This carried the RRL project number F-5100, and was supplied as a part of the airborne jamming system AN/APQ-20. Later, the Service nomenclature AN/APT-10 was assigned to the transmitter proper, which employed the Raytheon QK-44, 50-watt magnetron. Production of 56 AN/APQ-20 equipments were completed before the war ended.

In 1946, contracts were established by the Air Force for the development of an improved and expanded 150-watt jamming transmitter, extending in frequency through the L- and S-bands, to be known as the AN/APT-16 jamming transmitter. This project is a part of a program for the development of a series of jammers to cover the frequency range from 30 to 10,000 Mc. It has been supplemented, within these limits, by the AN/APT-13, a 150-watt, 70-150-Mc transmitter for jamming ground-to-air hf communications and uhf proximity-fuse signals; and outside the program range by the 400-watt AN/ALT-3 transmitter for the frequency range from 2 to 30 Mc. As a part of this same program, the AN/ALT-5, a 300-1000-Mc transmitter was added to bridge the gap between the AN/APT-6 and the AN/APT-16. These jammers are manually tuned, and adapted to be used in conjunction with a conventional search receiver and spectrum analyzer. The abilities to tune rapidly and to radiate increased powers are the improvements offered over World War II equipment.

On account of the requirement for automatic operation, a program was set up to develop a series of automatic single-frequency spot jammers covering the frequency range 30 to 10,000 Mc. The AN/ALQ-3 and the AN/ALG-7 were developed on a development-production contract to provide a rapid S- and X-band capability. These are airborne transmitters adapted to jam fire-control radars.

In 1947, the Bureau of Ships established a project for the development of a 1-kw, S-band shipborne jammer also provided with low-frequency coverage for communication jamming. The equipment, known as the AN/SLT-1 jamming transmitter, was placed in production. This transmitter covered the frequency ranges from 90 to 270 Mc and from 2460 to 3600 Mc at the 1-kw level.

A project for the development of the AN/TPQ-2 jamming system was established by the Signal Corps in January 1950. This system was designed for a 1-kw output in the frequency range from 9100 to 9600 Mc. It was intended primarily for use in making operational and engineering tests against airborne radar. Another Signal Corps jamming equipment was the AN/MLQ-2 equipment. This was a 1-kw jammer for the frequency range 2500-10,500 Mc. The complete system included an acquisition receiver, a setting-on receiver, and directional transmitting antennas operated in conjunction with an automatic target-tracking system. Each complete unit provided four channels, each one of which covered the entire frequency range. The actual system produced by Gilfillan Bros., Los Angeles, is the AN/MLQ-7 covering the frequency range 7200-10,500 Mc and providing the outputs described for the "planned" MLQ-2, which was not completed.

The Air Force 500-watt AN/ALT-1 transmitter is designed for the frequency range from 7500 to 11,000 Mc. This frequency range is divided into two bands, and is covered by means of two floating-drift-tube klystrons.

When the planning of the B-47 equipment was carried out, in 1952 and 1953, only manually tuned spot jammers were available; but an ECM operator was not planned for the B-47. As a result, a crash program for the provision of an unattended jamming capability for the B-47 was carried out. The existing jammers, the AN/ALT-3, the AN/ALT-6, the AN/APT-16, and the AN/APT-6 were utilized in this crash program. Since these jammers were provided with single-dial tuning they were converted to sweep jamming with the intention that they were to be preset to the center frequency of the band in which the radars were known to be operating and would then be electronically tuned over the radar frequency band. The AN/APT-6, the AN/ALT-3, and the AN/APT-16 were successfully converted. As the AN/ALT-3 could be swept only over that portion of its frequency band in which the one-quarter mode was used, it was discontinued. The swept versions of the AN/APT-6 and AN/ALT-3 as well as the AN/APT-16 were redesignated the AN/ALT-7, the AN/ALT-6A and AN/ALT-8, respectively. While slow sweep jamming is considered to be a poor technique today, large numbers of these jammers in dispersed airplanes were very effective against the radars that were in production prior to 1957. Better radar designs and ECCM fixes have provided a way to reduce the confusion effects caused by these jammers.

Projects under the sponsorship of the Bureau of Aeronautics include the AN/APQ-33 semiautomatic or manual search and jam system for the frequency range from 40 to 1600 Mc, employing conventional tubes; and the AN/ALT-2 jammer for the 8500-10,000 Mc range, employing a magnetron. Prototypes of the AN/ALT-2 were developed at NRL, and this transmitter is used in the AN/ALQ-23 automatic search and jamming system, which has

been adapted to the standard shipboard jamming system AN/SLQ-10.

Jammers for the low-frequency ranges, intended for communication and V-T fuze jamming include the AN/MRQ-2A, a 400-watt, 1.5-20 Mc communication jamming system (Signal Corps); the AN/TRT-2, a 100 watt, 75-200 Mc training jammer (Signal Corps), which is also provided with swept AM for fuzes; the AN/MRT-4 tunable jammer for operation at the 1-kw level in the frequency range from 75 to 1200 Mc (Signal Corps); and the AN/APT-13 swept jammer for the frequency range from 70 to 250 Mc (Signal Corps). ARDC also sponsored a variation of the AN/APT-13 equipment for the frequency range 70-1000 Mc, and a low-frequency communication jammer for the frequency range from 2 to 30 Mc designated as the AN/ALT-3 equipment. This development finally appeared as the AN/ALQ-14, a rapid search and jam system using a carcinotron.

The Bureau of Aeronautics established a project with the W. L. Maxson Company for the development of an automatic search and lock-on system at the 200-watt level, and for the frequency range 1000-11,000 Mc, which was designated as the AN/APT-20 equipment.

The AN/SLT-2 equipment was a 1-kw shipborne jammer for the frequency range from 15 to 1000 Mc developed by the General Electric Company under contract with the Bureau of Ships. A related equipment, the AN/SLT-3, was also a General Electric Company development, under BuShips sponsorship, with extended frequency range, to cover the spectrum from 950 to 10,300 Mc. This jammer, similar in many respects to the AN/MLQ-2, covered the required frequency range by means of 11 magnetrons.

Numerous studies have been made by contractors of all branches of the Services in the field of transponders for use against guided missiles and other types of targets. These studies included work on such equipment as the S-band and X-band "Owl" and the AN/ALQ-1 equipment, at the Airborne Instruments Laboratory.

A large amount of effort has been applied to various types of repeaters for use against proximity fuzes. The Electronic Defense Group at The University of Michigan accomplished a large share both of the research and of the development of experimental proximity-fuze repeaters. Some of the repeaters are designated for frequency ranges lying between 70 Mc and 300 Mc, and power outputs range from 0.1 watt to 10 watts. In general, these projects employ narrowband conventional tubes; some of them employ superregeneration; some are dual antenna repeaters; and some are single antenna repeaters. In this category are the AN/URQ-5, the AN/ARQ-16, the AN/ULQ-3, the AN/TLQ-2, the AN/TLQ(), and the AN/ULQ-1 equipments. Unlike these narrowband devices, the EDG at Michigan employed wideband, distributed amplifiers in the design of their successful V-T fuze repeater AN/

MLQ-8. This equipment employs one antenna, and a time-sharing transmitter and receiver. Further development of this equipment in tactical form was accomplished by Instruments for Industry. Two projects, ARDC Project R-112-132 and BuAer Project NL460-052, involved the study of traveling-wave amplifiers for use in broadband repeaters. Work along these lines was also done at the Stanford Electronics Research Laboratory under Joint-Service sponsorship.

Quite early in the postwar program it was realized that automatic jamming equipment must be provided for manned aircraft of the Air Force because it was not planned to provide such aircraft with ECM operators. In 1950 and 1951 study projects were established with the Sperry Gyroscope Company for investigating the feasibility of providing broadband coverage in semi-automatic and fully automatic search and jamming systems. These study projects resulted in the AN/ALQ-5 and the AN/ALQ-27 systems. A second technique was the AN/ALQ-6, in which a receiver and an FM carcinotron were combined to provide an automatic jammer capable of jamming, sequentially, all signals on the victim aircraft. Provisions were made to select the type of radars to be jammed by the use of prf and pulse-width discrimination. Another technique furnished multisignal jamming capabilities using a receiver to determine the frequencies present, control circuits to set up voltage-tunable magnetrons on these frequencies, and traveling-wave-tube amplifiers to raise the output signal to the proper level.

University programs for basic work in the ECM techniques area were established throughout the course of the postwar ECM program. Sizable programs were established at Stanford University, the Johns Hopkins University, and The University of Michigan, which have made, and are still making, significant contributions to the ECM program. For instance, The Johns Hopkins University work in infrared, during the earlier years, provided the knowledge to start a hardware program, when the need arose in 1956. At Stanford, work in the deception area was started on range-gate pull-off and inverse gain modulation as broadband multisignal radar track-breaking devices. Broadband multisignal repeaters were also investigated. Equipments, such as the AN/ALT-10, the AN/ALQ-11, the S-440, and the AN/ALQ-17 are the direct outgrowths of this effort. The Navy at NRL also sponsored a complementary repeater program for their ship and air research programs. Major technique contributions were also made by NRL in this area.

The Air Force has made important changes in policy with the objective of expediting the general procurement program. These changes include the weapons systems concept of development, involving the over-all contracting of an entire system, including the airframe, to a manufacturer; the establishment of a Quick Reaction Capability (QRC) for rapid hardware de-

velopment; and the planning and contracting of an infrared countermeasures
(IRCM) and radar absorbent (RAM) research and development program.
The use of infrared in air-to-air guided missiles and the added infrared radi-
ation due to missile launch started the need for infrared warning equipment
as well as infrared decoy devices.

2.2.2 Postwar Receivers

One of the outstanding accomplishments of the interim period since World
War II has been the successful development of the AN/APR-9 receiver,
under BuAer, which was placed in production. This receiver covers the fre-
quency range from 1000 to 10,750 Mc in four tuning units designated
TN-128, TN-129, TN-130, and TN-131. Its sensitivity ranges from —90
dbm, employing a narrow video bandwidth, to —73 dbm at the high end of
the frequency range, employing a wide video bandwidth. The development
work on this receiver was done by Airborne Instruments Laboratory, Air-
craft Radio Corporation, and Collins Radio Company.

Supplementing the AN/APR-9 receiver, the AN/APR-13 receiver was
developed by the Navy to provide coverage in the low-frequency range from
50 to 1100 Mc. This receiver is provided with five tuning heads, designated,
in order of ascending frequency, as the TN-200, the TN-178, TN-179, TN-
180, and TN-181 tuning units and has a sensitivity ranging from —87 dbm
at the low-frequency end of the range to —76 dbm at the high-frequency
end of the range, employing a wide video bandwidth.

The AN/BLR-1, AN/SLR-2 receiver is the shipborne counterpart of the
AN/APR-9 receiver, also including, however, the frequency ranges of the
AN/APR-13 receiver with the exception of the lowest frequency tuning head,
TN-200. Thus, the AN/BLR-1, AN/SLR-2 receiver provides frequency
coverage from 90 to 10,750 Mc, in eight tuning heads. The AN/BLR-1 re-
ceiver is intended for use on submarines, and the AN/SLR-2 receiver is
designed for installation on surface vessels and land stations. In both cases,
pulse analysis and direction finding systems are included.

At the end of World War II the Signal Corps Engineering Laboratory
developed the AN/TLR-1, the first ground-based intercept system to be em-
ployed by the Army. This equipment was designed by Polarad Electronics
Corporation, New York, and fabricated in quantity by Radio Corporation
of America, New York. The frequency range covered is 10 Mc to 12,100 Mc.
Three types of antennas are provided: omnidirectional, broadly directional,
and high gain. A signal analysis capability is included. Sensitivity ranges
from —120 dbm at the low-frequency end of the range to —90 dbm at the
high-frequency end of the range.

The U. S. Naval Research Laboratory pioneered in mechanical rapid-scan

microwave receiver development. Utilizing the basic mechanism of the AN/APR-9, it adapted a servo sweep system to provide scanning rates of roughly 1000 Mcs. The receiver also had the ability to present ten frequencies which were previously selected and return to these with great precision through the servo control system. Frequency was read on a Vaeder Root counter. The top four heads of the AN/APR-9 system were modified and included in the prototype system. Associated with this system development was a complete signal acquisition indicator utilizing the P-19 long-persistence screen and a time-frequency raster. Also a part of this system was a new indicator for signal analysis and display. A five-gun cathode-ray tube was utilized in which three traces were devoted to pulse analysis, one being an exponential sweep and two being two-decade log sweeps. These traces gave an indication of pulse width, prf, and antenna beam-pattern width. One sweep was used as a DF indication utilizing a rectangular sweep in which the top trace gave a downward deflection and the lower trace gave an upward deflection. The last trace is utilized for panoramic display. Production versions of this equipment resulted in the Navy shipboard set AN/WLR-1 and the airborne AN/ALQ-28 system. These systems have the same coverage as the AN/APR-9, 90 to 10,500 Mc, and the same order of sensitivity and selectivity.

The Signal Corps has also pioneered in the development of a mechanically tuned rapid-scan receiver known as the AN/TLR-9 and -10 receivers. The term "rapid-scan" is here used to identify a receiver in which the entire frequency range of a given tuning head is scanned in a period of time comparable with the duration of a radar "look," say 0.05 second. The AN/TLR-10 receiver is provided with five tuning heads and covers the frequency range from 60 to 1060 Mc. It is a superheterodyne receiver having sensitivities ranging from −101 dbm at the low frequency end of the range to −73 dbm at the high-frequency end of the range. The AN/TLR-4 extends the frequency range of the AN/TLR-10 from 1050 to 3600 Mc. The AN/TLR-7 is electronically tuned and employs saturable reactors. The AN/TLR-15 extends the frequency range of the AN/TLR-1. It covers 8 kMc to 41 kMc and utilizes backward-wave local oscillators.

Signal Corps projects for AN/TLR-3 and AN/TLR-17 receivers cover the extension of the AN/TLR-1 receiver. The AN/TLR-3 receiver has a frequency range from 12,000 to 18,000 Mc with one head, and the AN/TLR-17 receiver covers the range from 500 Kc to 12 Mc in eight tuning bands.

A somewhat similar airborne rapid-scan receiver, the AN/ARR-8A receiver, has also been placed in production by the Air Force. This receiver, a motor-driven rapid-scan superheterodyne, covers the frequency range from 70 to 960 Mc in four tuning ranges. Its sensitivity averages −80 dbm, and it employs an i-f bandwidth of 0.5 Mc.

In the rapid-scan receiver field, the Air Research and Development Center, ARDC, has sponsored two projects somewhat similar to the AN/ARR-8A receiver. These projects are the AN/ARR-1 receiver for the frequency range from 0.09 to 55 Mc, and the AN/ARR-35 receiver with a frequency range from 50 to 1030 Mc. These receivers differ from the AN/ARR-8A in that they are adapted for operation optionally as rapid-scan or as high-resolution receivers.

The AN/APR-14 receiver developed under ARDC sponsorship is a superheterodyne receiver, similar in some respects to the AN/APR-9 receiver, for the frequency range from 30 to 1125 Mc, using five tuning ranges. This receiver is intended primarily for setting jammers on frequency, and is also considerably more compact than the AN/APR-9.

The field of electronically tunable rapid-scan microwave receivers has been actively investigated, and work along these lines has resulted in operating equipment. The Electronics Research Laboratory, ERL, of Stanford University developed prototype models of the S-121 receiver covering the frequency range from 500 to 4000 Mc with three tuning heads. This receiver depends for its operation on the use of electronically tunable "dispersive" traveling-wave amplifiers, and scans over a 2:1 band in approximately .05 second. Frequency indication is provided by the display of the video signal on a scope, the time-base of which is synchronized with frequency. This is not a setting-on receiver, but does afford good frequency resolution, and assures almost perfect intercept probability on account of its extremely rapid scan rate.

In parallel with the Stanford project, and with the assistance of ERL, the W. L. Maxson Company has worked on a similar 2-4 kMc rapid-scan receiver known as the AN/ALR-2 receiver.

The Stanford S-151 receiver is an X-band converter designed to accept signals in the X-band and beat them down to the S-band range where they are fed into the input of the S-121 receiver. The S-152 receiver is an electronically tunable rapid-scan X-band superheterodyne with variable sweeping and acceptance bandwidth. The receiver is thus adaptable for high intercept probability with wide sweeping and acceptance bandwidth; and alternatively for high resolution, with narrow sweeping and acceptance bandwidths. The sensitivity is high in either case.

A considerable amount of activity is also going on in connection with wide-open receivers of various types. Some of these wide-open receivers are merely "zero catcher" receivers designed for early warning without any frequency indication. Others have multichannel operation, capable of indicating frequency with an accuracy of ± 15 percent. In this category falls the multichannel receiver of the Airborne Instruments Laboratory, in which

each channel employs a four-stage cavity filter for frequency discrimination.

Another type of wide-open receiver which is engaging considerable attention at the present time is a multiple-horn receiver being developed by deRosa at Federal Telecommunication Laboratories. Receivers of this type are discussed in connection with direction-finding systems in the next section.

2.2.5 Direction-Finding Systems

The program contains numerous projects for developing high-frequency direction-finding systems. These equipments are principally intended for use HF communication circuits.

The Navy at the U. S. Naval Research Laboratory has maintained one of the major research efforts in the field of hf direction-finding research. Since 1950 NRL has had an experimental wide-aperture direction-finder system in being and under study. Continuing studies have resulted in the development of suitable multicouplers for direction finding systems application. This has resulted in the development of a 40-element "Wullenweber" type array capable of simultaneously forming precise beams for very accurate direction finding, fixed beam antenna patterns every 9° of azimuth and less accurate steerable beams for intercept search purposes. Diverse instrumentation including several systems of combination intercept and DF functions have been developed. Twin-channel adaptations have also been included. Automatic bearing and read out to 0.1° bearing accuracy are also a part of the system developments.

The Navy has also sponsored direction-finding research at the University of Illinois. This has resulted in a 120-element "Wullenweber" array. Though the instrumentation connected with this antenna is quite elementary the studies produced have furthered the state of the art in appreciation of what aperture arrays wider than the 40-element system developed by NRL will produce operationally.

Also at the Naval Research Laboratory the Navy has sponsored a narrow-aperture direction-finding research program. Based on the AN/GRD-6 equipment, this program has resulted in studies of various antenna elements and their resultant accuracy as a part of the DF system. Major improvements in field modifications of the AN/GRD-6 have resulted. Automatic bearing readout systems, machine computation of fix systems and other similar operational concepts have resulted from these studies. The Signal Corps had a project with the Servo Corporation of America for the development of a doppler DF system, the AN/TRD-15.

The principal new development in airborne microwave DF is the AN/APD-4 equipment, and related equipment worked on by deRosa at Federal

Telecommunications Laboratories. This is a wide-open DF system which employs a multiplicity of directional horns, each with a crystal video receiver. The output of each receiver is connected to a delay line, and the outputs of the delay lines are scanned in timed sequence to produce a display of the over-all pattern on a cathode-ray tube. Although there is no precise frequency discrimination in the individual horns, a general indication of frequency can be obtained by examining the shape of the over-all pattern on the cathode-ray tube.

Two projects involving the improvement of the World War II AN/APA-17 direction scanning equipment are in the program. These are the AN/APA-92 direction-finding system for the frequency range from 60 to 11,000 Mc and the AN/APA-69 for the range 200-10,500 Mc. In this category is also the panoramic and DF indicator for the AN/APR-9 receiver, known as the IP-81/APA-69A indicator. A homing antenna system for the AN/APR-9, designated as the AN/ALA-4 equipment, is also being developed under BuAer sponsorship. This equipment covers the frequency range from 90 to 10,750 Mc and includes a commutation switch, signal comparator, indicator, and three sets of antennas. A. Alford has developed similar direction finding systems for the AN/BLR-1, AN/SLR-2 receivers now in production at Collins Radio Company. Another project is a dual-horn, servo-controlled DF antenna for the frequency range from 1000 to 11,000 Mc, developed as an improvement over the DBM direction-finding system.

2.2.4 Postwar Ferret Activities

After the war, the Air Force used some B-17's as short-range aircraft for ferret purposes. About the same time the Navy's postwar program began to take shape and PB-4Y2's, especially designed during the war for these purposes, with delivery at the close of the war, were organized in a ferret squadron. Later, the Air Force refitted a few B-29 aircraft for long-range ferret service. These flew some of the first global missions. The Navy's PB-4Y2's were still doing yeoman service with less extensive use in areas of strategic concern to the Navy. The Air Force later went into a program utilizing the B-36's.

The Navy followed the PB-4Y2 with the P4M. This aircraft served as a direct replacement for the PB-4Y2 and saw much service in the Atlantic and Pacific. Later, the Navy introduced an all-jet aircraft, the A3D-2Q. These aircraft are still flying and are typical ferret ECM-configured aircraft.

2.2.5 Postwar Supplementary Equipment and Studies

As in the case of the discussion of the World War II program, considerations of space prevent a detailed discussion of the many supplementary pro-

jects in the field of antennas, pulse-analysis equipment, and confusion and deception devices, which are a necessary part of the countermeasures research and development program. In general, postwar antenna projects involve work on improving the World War II antennas, on providing antennas having special patterns and characteristics, and on supplementary antenna equipment, such as connectors and switches. Pulse analyzers of various types are included, some of which are intended for use in automatic search and lock-on systems for the purpose of distinguishing between friendly and enemy signals. Work on window has continued, with particular emphasis on the development of improved reflecting materials and devices, and the development of various devices for dispensing the chaff.

The concept of saturation by the use of decoys was introduced in the postwar period. A research and development program to provide the technical knowledge for active and passive target enhancement was started, as well as development programs for both long- and short-range self-propelled decoys. The use of air- and ground-launched balloons as decoy devices was investigated, and developmental models were built and tested. From the initial simple devices the program developed into a long-range ground-launched decoy missile and short-range air-launched decoy missile, both with ECM equipment, to enhance the target return and a chaff capability. The long-range decoy also was given a limited jamming capability.

A new type of pulse-analyzer technique utilizing multigun cathode-ray tubes has been developed by the Naval Research Laboratory. In early 1947, NRL began developing multigun cathode-ray tubes. This development established a new generation of cathode-ray tubes for military application in this country, resulting in joint Army-Navy specifications for three- and five-gun, flat-faced, rectangular screen tubes having excellent performance characteristics of spot size, deflection sensitivity, and trace location. In 1948 the first experimental multigun signal analyzers were installed in a new ship configuration aboard the U.S.S. *Columbus*. This equipment was later put into production, and was known as the AN/SLA-1 pulse analyzer. It utilized five sweeps of different lengths, the first being 0 to 5 μsec, the second 0 to 50 μsec, the third 0 to 1000 μsec, the fourth 0 to 5000 μsec, and the fifth 0 to 50,000 μsec. An airborne version of this same equipment was the AN/APA-74. Its characteristics were identical to those of the AN/SLA-1 except it was packaged for airborne application. The Army's contribution to this field is the AN/ULA-2 developed by USASRDL.

Another three-gun airborne pulse analyzer, the AN/ALA-3, utilized exponential sweep for its displays. Further developments in integrated display units for pulse analysis, direction finding, and panoramiscope presentations were developed for the rapid-scan equipments which later became the AN/

WLR-1 and the AN/ALQ-28. Here the pulse-analysis scope utilized one exponential sweep 0 to 10 μsec long and 2 two-decade log sweeps, one 10 to 1000 μsec and the second 1000 to 100,000 μsec long.

Investigation on fundamental circuitry was done at NRL for demodulating all types of pulse modulation. These units incorporated features for pulse width, pulse length, pulse period demodulation, or any combination of the above. Also included was circuitry for generating a synchronizing pulse when no sync pulse was transmitted with these systems. Because of the highly specialized nature of these equipments, only a few models of these systems were developed for very special applications. The techniques developed were adopted, however, for use by other countermeasures groups in missile telemetering demodulation in many specialized applications.

2.2.6 Current Developments in Vacuum Tubes

The subject of vacuum tubes is of prime importance to the counter-measures research and development program, and some consideration of this matter is therefore required. It is, however, difficult to give a complete picture of the vacuum-tube situation without listing available and developed tubes in tabular form. Numerous tabulations of vacuum-tube developments have already been made, or are in the process of being made. The following discussion will therefore be confined to a summary of the vacuum-tube program.

Conventional tubes are now available or in the late development stages, for frequency ranges as high as 1000 Mc, at the 6-kw level.

The CW magnetron program which was established during World War II continues. Magnetrons for the frequency range from 90 to 1200 Mc and for a power output level of approximately 150 watts are available in limited production. Development has been completed on several CW magnetrons at the 150-watt level for S-band and X-band frequencies.

In the 1-kw range, difficulties are still being experienced in obtaining satisfactory CW production tubes in sections of the frequency range required by the countermeasures program.

In the klystron field, high-powered klystron amplifiers in the uhf region, intended for uhf television applications, are available. Studies are being carried out to determine whether floating drift-tube klystrons have sufficient tuning range, power output, and efficiency to serve as a useful interim substitute for the 1-kw magnetron which are being delayed. Klystron amplifiers capable of pulsed outputs in the megawatt range at S-band have been successfully produced at Stanford and General Electric, but these tubes are not generally satisfactory for countermeasures applications.

Several varieties of Resnatron oscillators and amplifiers capable of high-

efficiency CW power outputs as high as 50 kw h ve been produced in the form of laboratory models. None of these tubes has, however, been sealed off, and there appears to be no present prospect that sealed-off Resnatron amplifiers or oscillators having a satisfactory life will be available in the future.

A considerable amount of work is now being done on broadband traveling-wave amplifiers and dispersive traveling-wave amplifiers. This includes work at Stanford on dispersive amplifiers (for the S-121, AN/ALR-2 receiver), low-noise receiving traveling-wave tubes operating as high as X-band, 1-watt S-band amplifiers, 150-Mc 40-watt amplifiers and 100-watt S-band and X-band amplifiers. Pulsed helix-type amplifiers for power outputs as high as 1-kw have been completed at Stanford for S-band and Kw-band applications.

At the present time electronically tunable oscillators and amplifiers, such as backward-wave oscillators and carcinotrons, are being emphasized in the program. It is probable that tubes in these categories will be found suitable for operation at frequencies over the entire spectrum from 100 Mc to the millimeter region, and at power levels ranging from milliwatts to kilowatts. Investigations of backward-wave oscillators at Stanford and Bell Telephone Laboratories have now resulted in the development of tubes for various frequencies by a number of companies.

"Carcinotron" is the name given to a tube of this type by its inventor, M. Warnecke of the Compagnie Générale de Telégraphie Sans Fil, Paris, France. There are two types of carcinotron; a type M carcinotron which employs crossed electric and magnetic fields, and a type O carcinotron, which is similar to a backward-wave oscillator.

The General Electric Company and The University of Michigan are working on an equally important tube, the voltage-tunable magnetron, which, like the backward-wave oscillator, promises to have important countermeasures applications.

3

Perspective of ECM in Modern Warfare

J. A. BOYD, D. D. KING, G. L. LANSMAN

3.1 Introduction

The primary function of electronic countermeasures is to destroy the utility of enemy communications, weapons, and surveillance devices. The utility of these devices can be destroyed in varying degrees by denying information or by providing erroneous information. The eventual result of the use of electronic countermeasures may be physical destruction, as in the predeto   n of a missile warhead, or perhaps denial of use, as in the confusion oi a : acking radar system.

The bac. function of electronic countermeasures has remained the same since its inception, but manifold applications have been made over the years. Because of the multiplicity of applications, it is helpful to consider various aspects of electronic countermeasures in light of the geographical basis of the several phases of warfare. Separate consideration of these phases shows the variations (in the countermeasures) caused by the parameters space, time, and environment. The divisions of this chapter are based on geographical environment: (1) airborne ECM, (2) ECM in surface-type naval operation, (3) ECM in air defense, (4) ECM and ground operations, (5) ECM and underwater operations, and (6) ECM in space. In the subsequent chapters, the material has been organized on the basis of technical significance. The reader will recognize that many techniques are used in all six of the areas of warfare mentioned above. However, the various parameters and special problems make the geographical breakdown useful.

3.2 Airborne ECM Systems

3.2.1 General ECM Functions

Electronic countermeasures perform two general functions in aircraft: (1) defense and (2) reconnaissance. The electronic defense of an aircraft includes active and passive means for defeating attacking weapons relying on electronic detection, aiming, fuzing, or guidance methods. The reconnaissance or data-gathering function provides information on the types of signals and their sources for either tactical or technical uses. In a given aircraft, both defensive and reconnaissance functions are usually combined in varying degree to suit the mission requirements. An aircraft primarily intended for reconnaissance nevertheless requires defensive protection. Similarly, in a mission with no data-gathering functions, some information on the signal environment is needed to operate the defensive equipment. Since the aircraft must survive to accomplish its mission, sufficient defensive strength to assure the necessary survival probability is the primary ECM consideration. Attack on an aircraft penetrating a defended area is controlled by early warning (EW), surveillance, airborne intercept (AI), and various tracking radars. Communication, guidance, and fuzing signals may also be essential to the attacking weapons. The effectiveness of various electronic countermeasures in protecting aircraft penetrating enemy defenses can be very great. Conversely, improper use of countermeasures can destroy the often vital element of surprise and assist intercepting weapons in finding their targets.

Countermeasures available for aircraft defense and airborne reconnaissance equipment are considered separately in the first sections to follow, and finally in combination to form systems.

3.2.2 Countermeasures against Early Warning (EW) Radars

Detection must precede any attack; hence, avoiding detection is the earliest countermeasure at the disposal of the aircraft defense. Early warning (EW) radars, on the surface or airborne, first detect approaching aircraft. Therefore, evading detection by such radars constitutes the initial step in aircraft defense.

The maximum detection range r of a given radar is proportional to $\sqrt[4]{\sigma(\theta, \phi)}$, where $\sigma(\theta, \phi)$ is the echoing area for the particular aspect (θ, ϕ) presented to the radar; to halve the range, r, $\sigma(\theta, \phi)$ must be reduced 12 db. Thus, even modest delays in detection can only be purchased by order-of-magnitude reductions in aircraft reflections. These reflections are critically dependent on the angles (θ, ϕ), and may vary as much as or more than ± 20 db from maximum to minimum. If detection is to be escaped during

flight over an area or perimeter defended by radars separated by a spacing R, the detection range must be reduced to $r < R/2$. The problem of reducing $\sigma(\theta, \phi)$ sufficiently to accomplish this over the probable ranges of frequency and angle is severe. The resultant penalty in aircraft performance with techniques currently available becomes excessive, even though the only counter-countermeasures are to make more sensitive radars or to space them more closely.

Hiding the aircraft in a cloak of radar invisibility, although potentially most effective, therefore appears unfeasible against a well-designed defensive perimeter. Failing this, the next best choice might be to conceal one or more aircraft behind a cloak which is itself visible, but obscures the nature and location of the true targets. This alerts the ground defense that penetration of its domain is under way or in preparation, but withholds vital information on target locations and characteristics. Such a cloak may be produced by either active or passive means.

If a substantial volume surrounding the aircraft is filled with reflecting elements, i.e., with chaff, the aircraft echo is concealed in all directions by the profusion of spurious echoes. Such a corridor must be sowed. However, the sowing vehicles are also partially concealed by the reflectors they dispense; their presence can only be deduced as being near the end of the corridor.

The sowing of a corridor or group of corridors is an elaborate operation which reduces the time and space available for a given mission. Depending on the type of weapon deployed against a penetration mission, this may or may not have adverse effects on survival probability. Generally, the closer the weapon control from the ground, the more effective the corridor technique becomes. Against MTI* equipped radars, the corridor technique increases the background clutter, and hence reduces the usable range. In areas with large changes in wind velocity at various levels (wind shear), MTI radars are particularly vulnerable. In general, the presence of chaff severely hampers the EW radars, and various types of MTI offer only partial protection to the radars.

The radar echo from an aircraft also may be masked by jamming signals. These simply override the true return in the receiver input circuits. The resulting strobe display pattern is easily recognized, and can usually be made to yield azimuth information. The absence of range data in the strobes prevents direct target location, since triangulation from multiple radars produces many spurious solutions.

A barrage noise jammer raises the noise level over a broad frequency spectrum, including the receiver acceptance band, to a level determined by

*Moving-target indication.

the jamming signal strength at the receiving antenna. This is equivalent to increasing the noise figure of the receiver, and deteriorates the radar performance in the same manner. When pushed far enough, this "brute force" or thermal method can render ineffective active radar operations by raising the background noise to intolerable levels. Negation of warning radars by barrage jamming is a relatively sure method, but requires massive power to cover much of the available spectrum. The barrage method also invites the use of counter-countermeasures, including special triangulation procedures and correlation receivers for locating individual jammers.

The broad-scale countermeasures against warning radars discussed so far make little use of specific properties of the radar. By concentrating the effort on a particular aspect, economy in jamming can be achieved. Spectral concentration involves placing the jammer power only in bands occupied by enemy radars. This results in orders of magnitude greater spectral efficiency, i.e., many times the useful jamming signal per watt of transmitter power. A narrowband or spot jammer must be set on the radar operating frequency. Automatic set-on circuitry to do this is complex, and often easily defeated by too many signals, or agile frequency changing by the radar. Preset or manual set-on for spot jamming is, of course, even more easily countered in the same fashion.

Most efficient in principle is the concentration in both frequency and time achieved by the deception repeater or transponder. Here, pulses in addition to the true echo are transmitted to produce many false returns. The false targets may simply appear rather randomly to confuse the radar, or they may be carefully programmed to have realistic courses. Both the simpler confusion device and the complex programmed repeater usually enter the radar antenna through its minor lobes, and can be countered by the proper rejection circuitry. In the absence of such circuitry, changes in target size and scintillation may enable an experienced operator to separate true and false targets.

For every target on the display, the radar operator must make a decision: is it a true target or false? If a true target is considered false there will be no interception, hence a minimum of incorrect decisions by the radar operator is mandatory. Alternatively, false targets accepted as true dilute the available intercepting forces. If the total number of targets accepted as true exceeds the capacities of the intercepting forces available, saturation of the ground defenses may cause a breakdown in the normal tactics and prevent proper assignment of weapons to bomber targets. False targets produced by electronic means can be identified with sufficiently elaborate circuitry or possibly by carefully trained operators. A relatively sure method of providing the requisite number of targets is to use small decoy aircraft in large

numbers interspersed with the bomber force. Within limits, small and cheap decoys can be provided with enhanced echoing power to render them indistinguishable from bombers for the radar. Rather elaborate changes in the early warning radar system may be required to circumvent the decoy problem. One such change might be the use of bistatic radars to identify the highly directive scattering pattern of corner reflectors used in enhancing the decoy echo.

3.2.2 Countermeasures against Airborne Intercept (AI) and Tracking Radars

Reduction in the effective range of airborne intercept (AI) radars has an effect similar to a proportional reduction in early warning range. In both cases, less distance from the target is needed; in the airborne instance this means not only more radars, but also more interceptor aircraft. Complications such as MTI are also greater penalties in airborne equipment. Denial of range information by barrage or spot jammers is of less importance since azimuth data may suffice for the interceptor to close the range. On the other hand, a false target accepted as true for only a short time may place the interceptor or missile on a course from which the target can no longer be reached with available fuel supplies. The possibility of homing on barrage jammers must also be considered. Intercept operations can be facilitated by strong barrage signals, if the interceptor is equipped to home on them. However, the homing capability must always be auxiliary to some other acquisition and tracking system. In a sense, homing is, therefore, a second-line threat.

As the range is closed, the intercepting aircraft or missile switches from the search to the tracking mode of its AI radar. At this late point in the engagement various schemes for breaking lock in the tracker remain. A tracking radar provides accurate range and azimuth data to direct weapons to within lethal radii of the target. Therefore, an aircraft being tracked by one or more hostile radars is probably in acute danger. Aircraft defense systems assign correspondingly high priority to countering a tracker once it locks on its target.

High antenna gain and range gates concentrate the effective radar power in space and time. This makes it difficult to cope with the radar once it has locked on target. The time available for dislodging the tracking radar is also short. In return, the time for reacquisition, once tracking is broken, is even shorter. At close ranges, any protracted break or error in the tracking data can result in a complete loss of the target.

For many probable angles of illumination by a tracking radar, it is possible to sharply reduce the aircraft radar cross-section by suitable shaping

of metal surfaces and the use of absorbing materials. The resulting reduction in echo facilitates the function of jammers. Overpowering a tracking radar by noise jamming is difficult unless aided in this manner because of the power requirements. However, several schemes successfully utilize specific characteristics of the tracker. These include chaff, range-gate-grabbers and inverse gain repeaters. In each case, the radar is pulled off the true target and forced to search again to find it. In each case it is also possible to protect the radar against the countermeasure to a considerable degree, although not completely. This protection is purchased, however, by complexity which may be unacceptable in very compact systems.

The use of phase comparison (interferometer) seekers in missiles, and various other nonscanning-type trackers poses special problems in countermeasures. Noise jamming can destroy range data, but the radar is generally able to track the noise source without difficulty. However, the natural limitation of tracking in the form of target scintillation or glint can be enhanced to jam phase comparators; by utilizing several widely spaced antennas, a false phase front can be generated which produces a small angular error in tracking. Changing the tilt of this false phase front introduces an oscillating angular error into the tracking antenna. The dynamics of the tracking servo system can then be utilized to produce large angular excursions. The locus of these excursions can, by proper location of a least three sources, be made to lie far from the target at all times.

A related tactic is feasible against infrared seekers which home on the hot surface of the aircraft. If the tracker can be made to follow a bright decoy flare for a brief interval, the large angular error introduced makes it difficult for the seeker to reacquire the true target in the time available. Both the limited field of view and the servo dynamics of infrared seekers offer possible fields for countermeasures.

3.2.4 Communication, Guidance, and Fuse Countermeasures

Unless the acquisition range of missile seekers or AI radars is very large, additional guidance from the ground is required to achieve contact between attacking weapons and the aircraft under attack. For manned interceptors, guidance may be by voice command, or by data link. In automatic systems, data is transmitted over a special link, which may consist of coded pulses from the ground tracking or surveillance radar. This intermediate stage of control or midcourse guidance occupies a special position in countermeasures for aircraft defense.

Where the information is transmitted in uncoded form by relatively low-power omnidirectional ground stations, the link can be broken rather easily. The result is that interceptors must follow preassigned courses. Such "broadcast" or "single-vector" control materially reduces the probability of inter-

cept. However, the jammer may serve as a beacon if the attacking aircraft is equipped with homing equipment.

By communicating over a directional channel, such as a tracking radar, or by suitably encoding the data, the ground station can secure reliable contact with an interceptor or missile. Thus, the ground operator can confine his transmissions in time or frequency, as well as in solid angle, so as to place the airborne jammer at a great disadvantage.

Because of the widely differing results which can be realized, jamming of communication and guidance links demands particular attention. The rather general principles applying to various types of radar countermeasures must be replaced by specific rules applying to each particular situation. For example, airborne communications jammers have been used with success against enemy tactical ground communications, such as between tanks.

Jamming the fuze of a missile by prefiring or inactivating is the last recourse available in aircraft defense. Certain types of active fuzes are susceptible to jamming. However, a fuze is inherently a low-information device; it need only note its arrival on a predetermined surface about the target. The antennas and circuits to achieve this can be specialized to narrow beamwidths and narrow bandwidths, respectively, with resulting high countermeasures resistance. Because of the low probability of encountering both an active jammable fuze, and of being able to cause its malfunction, fuze jamming occupies a minor place in aircraft defense.

3.2.5 Electromagnetic Reconnaissance

Reconnaissance or "Ferret" missions are flown to acquire information on the electromagnetic radiation from enemy sources. Two broad classes of information are generally sought, namely, detailed technical intelligence information giving insight on new enemy equipments, and "order-of-battle" information for immediate tactical evaluation. Actually, there is considerable overlap between the two, notably because of the desire for detailed information on particular enemy equipments for tactical purposes. The technical intelligence mission seeks maximum information, and hence takes as much time, space, and equipment as is permissible. A large aircraft loaded with electronic equipment, and supported by extensive ground-analysis apparatus, may be needed for such a mission. Some "order-of-battle" information may be gathered by simpler apparatus and is often flown as part of another mission. On the other hand, for complete coverage of all signals, the order-of-battle data desired approaches that for technical intelligence.

Three principal types of information are present in radio signals: (1) frequency, (2) modulation, and (3) source location. To these must be added signal strength and the all-important criterion of presence or absence. The accuracy required in distinguishing the type and nature of information deter-

mines the type of equipment used, and hence the nature of the mission.

A high-flying aircraft is exposed to a tremendous number of signals. The largest and perhaps most important class, radar signals, may number in the thousands, with fewer communication, navigation, identification, guidance, and jamming signals. To cope with this mass of information, some means of selection is required. The particular choice of selection process determines to a large extent the capabilities of the ECM system employed.

The simplest type of system is the activity indicator. Here the selection is based on amplitude only: an omnidirectional antenna (in the horizontal plane) and a broad band receiver are coupled to a threshold device. The primary use of this system is as a warning device and actuator of aircraft defense systems. Signal amplitude is a measure of range, and hence of danger to the aircraft. By setting the threshold at a high level, only the strongest, and hence potentially most dangerous, signals are accepted. However, none of these are missed: the intercept probability for dangerous signals is unity.

A more complex receiver is included in automatic jamming systems. Here, the warning receiver is expanded to determine whether the signal is indeed a dangerous type. If the answer is affirmative, the receiver output actuates defensive measures.

Simple direct-detection receivers provide adequate sensitivity for detecting the main beam of radars. Minor lobes from radars and also most other signals are below the noise level except at close ranges. The number of signals the system must accommodate is much reduced by the lower sensitivity, but nevertheless hundreds of pulses per second may be received. Some selection in frequency can be accomplished without sacrifice in intercept probability by parallel filters. Multiple directive antennas similarly provide selection in azimuth. Pulse coincidence circuits between frequency and angle selection permit identification of pulses by both criteria. Selected pulses sorted in carrier frequency and azimuth angle may be stored directly on tape or film. Additional analysis may also be performed at the time of reception to categorize the types of pulses, and thereby further reduce the data remaining for final study.

A broadband nonscanning system of the general type described provides data on the number and location of all radars illuminating the aircraft. This information satisfies the most urgent tactical requirements. Some additional information is provided on the nature of the signals, and hence on the probable types of radars. This is of value for both tactical and technical evaluation purposes. The entire system can generally be made compact, and suitable for installation in smaller vehicles or as supplementary equipment in larger aircraft.

A basic problem of electronic reconnaissance lies in the huge amount of

information available to be extracted from the signals within the range of a sensitive receiver. A relatively simple, sensitive, and selective receiver is the superheterodyne. However, its high sensitivity is purchased at the price of intercept probability, as can be seen from the simple considerations below.

Receiver noise places a lower limit on the detectable signal strength. For a given false-alarm probability, the minimum detectable signal strength is proportional to the effective receiver bandwidth and inversely proportional to the antenna gain. Evidently the most sensitive systems are highly selective in both frequency and solid angle. Such systems often require a relatively long time to cover a broad frequency spectrum and solid angle. Therefore, their probability of intercepting signals periodic in time can become small. In the case of signals from scanning radars, the radar scan periodicity is effectively eliminated by sensitive receivers which receive signals from the minor lobes of the radar antenna. For other types of periodicity, such as intermittent operation, added sensitivity cannot compensate for the slow rate of coverage or slow receiver scan.

Given indefinite time, the single channel superheterodyne can locate and identify all signals in sequence. In the transient situation of an aircraft flying rapidly over areas containing periodic sources, only a sample can be obtained. This is particularly true if a swept-frequency receiver is coupled to a scanning direction finder.

Alternative systems involve many parallel channels with broadband traveling-wave amplification. Supplementary tuning is then required to determine the frequency. Systems with very fast sweeps, with separate direction finders, with many parallel channels, and various other combinations have been proposed to provide greater resolution.

For systems intercepting many signals, the analysis of the data is a difficult task. Video recording and subsequent ground-based analysis affords the most complete results. Limited airborne analysis, followed by recording of selected signals, perhaps at an operator's discretion, facilitates handling the mass of signals, and permits concentrating on the important ones.

3.2.6 ECM Systems for Aircraft

The link between reconnaissance and aircraft defense is the warning system. This may consist of special receivers and infrared detectors which actuate the defense equipment. The presence of certain types of signals above a prescribed threshold level suffices to determine the need for chaff ejection or barrage-jamming operations. More complicated jammers carry their own integrated receiving systems which provide the necessary input data such as frequency and modulation for correct jammer operations.

In a broader sense, the information-gathering (sensing) function should

be coupled to the defensive equipment (action) through a computer. Here, the input data on the instantaneous situation is processed and an optimum over-all defensive procedure calculated according to predetermined criteria. This permits coordinating all defensive means, including aircraft maneuvers, defensive fire, and the various electronic countermeasures. The very short duration of an engagement between supersonic vehicles requires such an electronic programming of the defense.

The number of potentially useful ECM functions, both defensive and information seeking, is so large that no single aircraft can perform all of them. A compromise must therefore be made in choosing the ECM capability for a particular aircraft and mission. Space, weight, and interference with other equipment are the principal limiting factors. The space for electronic equipment in aircraft is limited, and often unfavorably located. This is particularly true in the case of controls and antennas. Controls can be eliminated to a large extent by proper design and automatic operation. Unfortunately, antennas for many important functions may require the same space on the aircraft surface. Multiplexing is a partial solution to this problem, but often limits the availability of equipment.

Interaction between various equipments that can be carried simultaneously usually occurs via power supply or antenna. Operating all equipment at once may overload the generator capacity. Antenna interaction between receivers and radars can often be blanked out. Jammers can be treated in a similar manner by blanking certain friendly frequency regions in the jamming spectrum.

Weight penalties are measured directly in reduced aircraft performance. Each pound of defensive ECM equipment must therefore be evaluated in terms of its contribution to the probable success of the mission. Reconnaissance-type equipment is perhaps best evaluated in terms of the number of aircraft required to produce the desired information.

3.3 ECM in Surface-Type Naval Operations

3.3.1 Surface ECM Functions

Surface warfare at sea is inseparably linked to air operations. As a result, shipborne systems are closely integrated with airborne equipment. Reconnaissance, for example, proceeds best from the air, and an exclusively shipborne intercept system would have limited coverage. Similarly, no defense based exclusively on shipboard equipment is probable. The over-all scope of fleet defense therefore is broader than either surface or air operations taken individually. The role of ECM in surface-type operations is correspondingly complex. Integrated systems for surface ECM have not

evolved to parallel airborne systems. Nevertheless, individual functions have been highly developed, and these will be mentioned.

Targets and time scale in sea warfare differ significantly from those for airborne systems. Some ships have very large radar cross sections, but submarines can reduce their echo to zero. For low angles of attack, the horizon and sea clutter hamper radars. However, from high altitudes, the uniform surface of the sea provides a nearly perfect background for detecting surface vessels. Ships move slowly when compared with aircraft, and the duration of engagements is therefore longer. The interval during which reconnaissance data can be gathered is also longer, except in the case of submersible targets.

To grapple with the wide range of conditions to be encountered, seaborne ECM equipment must have high performance and flexibility. Fortunately, the space and weight requirements for antennas and equipment are somewhat less severe than in airborne applications. The equipments and tactics for surface ECM at sea are discussed below. The general topics considered show close parallelism to the airborne situation. However, the details and emphasis differ materially.

Reducing radar visibility or echo, and providing false echoes are both potentially effective countermeasures against radars. At sea, the application of nonreflecting materials is sometimes hopeless, sometimes highly effective. Thus, the echoing area of a 1000-foot vessel might be many acres. A reduction sufficient to affect the detection range seems out of reach in this case. In contrast, a submarine periscope and breathing tube can be effectively concealed by an absorbing coat. The combination of a small echo and the surrounding sea clutter makes a concealed periscope very difficult to detect by radar. In ships of small and intermediate size, the possibility exists of reducing or focusing the return by absorbers or flat surfaces.

The smoke screen familiar to naval operations in past years has a radar counterpart in chaff clouds. In surface warfare, such clouds are difficult to sow and control because of wind conditions and settling rates. The use of discrete false targets or decoys on or above the surface of the sea is a more effective way of concealing true targets. A host of false returns from decoys provides almost complete protection if decoy and target are indistinguishable to the radar.

A noise-modulated jamming signal obscures range information. However, the azimuth of the jammer can usually be determined. In air-to-air engagements, the relative speeds and ranges of attacker and victim aircraft are such that range data is needed to plot the required intercept course. Homing on an azimuth heading is effective only when the attacker has a large speed advantage. In an air-to-surface engagement, this speed advantage exists. Hence noise jammers offer limited protection against air attack on an iso-

lated ship. On the other hand, a powerful noise jammer can obscure the number of targets in its vicinity by saturating the victim radar receiver. In the case of surface search radars, denial of range information and obscuring of the number of targets materially reduces the value of the radar.

The noise jammer by its own radiation provides an azimuth heading for the attacker. This vulnerable feature is eliminated in a properly operated deception jammer. This makes deception equipment attractive for defending individual ships. The space available on shipboard also makes feasible more complicated techniques than might be usable in aircraft. Azimuth repeaters against search and bombing radars, as well as various range and angle repeaters against trackers, therefore are potentially effective components of ship defense. Relatively simple transponder- and repeater-type devices may be used on decoys to enhance the return and thereby improve the target-deception effect.

Active jamming against enemy communications has limited usefulness in operations involving surface vessels. The presence of a strong jamming signal may provide long-range homing information which assists the attacker more than he is hindered by the jamming of his communications. After direct contact has been established between the attacker and target, communications are, of course, no longer necessary.

Long-range missiles employ a variety of guidance and navigation signals which are susceptible to countermeasures. With sufficient information on codes and frequencies, the missile can be denied vital information and possibly can be given false data. Shorter-range missiles with line-of-sight guidance by beam-rider or command are less vulnerable, largely because of concentrated antenna patterns.

The final stages of a missile attack may be controlled by self-contained seekers and fuzes. Here, the surface vessel enjoys some natural protection from the sea return, and also offers a large platform for antennas, jamming equipment, and intelligent operators.

3.3.2 Shipboard Reconnaissance

The signal sources within reach of shipborne intercept equipment differ significantly from those accessible to aircraft. In the first place, fewer signals are encountered, partly because the ocean is sparsely populated relative to defended land areas, and partly because of the smaller area enclosed by the horizon at sea level. A second important point of difference is relative motion. Sources on the ground are mostly stationary, with the exception of a few in slow motion relative to the observing aircraft. At sea, most important intercepts are those from aircraft or other ships and submarines. Exceptions might be long-range guidance and communications signals at relatively low frequencies, as well as signals received from tropospheric scatter.

The motion of the intercepted source is an important item of information, and requires repeated position determinations. The intercept equipment must also be able to cope with intermittent operations, as in the case of a submarine surfacing for a brief period. The observation time may therefore be short, even though relative motion is slower than in air operations. Bearing information and signal identification is, of course, of vital importance to defense against attack from the air or the sea.

The availability of space and personnel on shipboard as well as the reduced signal density make manual supervision of intercept equipment possible. This permits greater flexibility in data processing and interpretation.

3.4 ECM in Air Defense

3.4.1 Opportunities for ECM in Air Defense

The role of countermeasures in air defense is confined to two rather diverse areas. The first of these concerns action against navigation and bombing radars in attacking bombers. The second properly is a counter-countermeasure, since it attempts to make use of the signals from radar jammers in the attacking aircraft. This technique of passive detection and tracking of jammers does not resemble the antijamming "fixes" applied to radars. Therefore, even though radars are involved in the passive systems, the techniques involved are generally grouped with countermeasures, and not with counter-countermeasures or antijamming devices. Since neither navigation and bombing radars nor jammers may be used by the attacker, ECM in air defense is strictly an activity of opportunity. The importance of such an activity depends on the probability of need, and on the potential effectiveness when used. Some judgments of these factors are necessarily included in the discussion of the two countermeasures areas mentioned, even though the primary emphasis is intended to be on ECM techniques.

3.4.2 Countering Bombing and Navigation Radar

Inertial and astronomical reference frames are sufficient to provide guidance to within a specified probable error. Where this error meets tactical requirements, radar aids are not required and countermeasures equipment is superfluous. However, a dead-reckoning approach to the target may not be adequate where a precise delivery point is demanded. In addition to the inherent errors in map measurements and in the navigation system, extraneous factors such as wind velocity and changes in target position contribute to the total miss distance. The use of radar aids to supplement the navigation system can then be expected. To locate a stationary target, only occasional fixes are needed for course correction. Intermittent radar operation

suffices to provide such corrections. Furthermore, the radar reference point
need not be close to the target if suitably offset references have been
selected. For moving targets such as ships at sea, the offset technique
evidently does not apply.

The intermittent signal of a bombing and navigation radar is inherently
difficult to counter. Also, it may not be desirable to merely deny the navi-
gation information sought by the radar at one point. In this case, the
bombardment can still proceed with reduced accuracy. To be entirely suc-
cessful, the countermeasures should provide false navigation data which
will maximize the miss distance at the target. In addition, the presence of
several radars in the bombardment force is to be expected. Therefore, the
ECM system must be capable of dealing with multiple signals from dif-
ferent locations. To perform such a task, the ECM system may include two
interrelated functions. These are: (1) obscuring the real targets and (2)
providing false target information. Several methods for accomplishing these
aims are available.

Perhaps the most obvious method of obscuring target return is to reduce
the echo from the salient geographical features noticeable to the navigation
and bombing radar. However, the large-scale changes in physical contours
and liberal use of absorbing material required make this approach practicable
on a very limited scale only. Another method of changing the radar topog-
raphy is to provide false returns that are sufficiently large and numerous
to mask the existing features of the area. Both reflectors and repeater jam-
mers are capable of producing suitable false echoes. Finally, it is possible to
raise the background noise level by means of numerous noise jammers.
This type of interference provides a threshold level below which the weaker
echoes are masked. A carefully designed combination of false targets from
reflectors and repeater jammers, supported if necessary by noise jammers,
is capable of drastically altering the radar appearance of an area. In some
cases, the proper location of only a few small jammers has produced good
results.

The concept of a fixed area defense against many attackers does not
apply in some cases. A ship at sea or an isolated military installation might
be more efficiently defended by countermeasures aimed at totally disabling
the bombing radar in the vicinity of the target. A powerful jammer with a
directive antenna may concentrate sufficient power on the bombing radar
to saturate the receiver through the minor lobes of the radar antenna. If
radar data are the only guide available to the bombardier, he may then be
obliged to attack the jammer for lack of a better target.

3.4.3 Passive Detection and Tracking
There are two broad classes of radar activity in air-defense operations.

The first of these concerns detection, and is handled by long-range scanning radars. The second is target tracking, which is done by tracking radars of suitable precision. Both of these radar functions are susceptible to jamming. In each case, the effect of jamming is to reduce the maximum target distance at which radar range data become available. Since range information is essential to both detection and tracking functions, some means of using the jamming signal to provide range information is needed. The techniques available for this are now considered in turn for the detection function and for the tracking function.

Detection of targets by a scanning (early warning) radar includes identification of a target in space and determination of its course. For this purpose, the radar output is placed on a PPI display, on which the targets appear as dots. When a noise jammer of sufficient power is activated on the target, the dot becomes a strobe line. As the jamming signal grows larger relative to the target echo, the radar gain must be reduced and the side lobes suppressed to maintain a narrow strobe. A basic characteristic of this situation is that azimuth data remain available. In a defense net involving many radars, the intersection of strobes from several radars defines a target location. Unfortunately, when many jamming targets are present, there are many false intersections or ghosts. Thus, when two jamming aircraft are received at two radar sites, two azimuth angles are obtained at each site. From these four possible positions can be calculated, of which two are correct and two are ghosts. In general, n jamming aircraft produce n^2 positions of which $n^2 - n$ are ghosts. When three radar sites are used, it is possible to resolve the targets in principle. In practice, this requires a special display and a skilled operator, and is reliable only when the situation develops slowly and the number of intersections is not too large. With increasing numbers of jammers, the elimination of ghosts becomes more difficult. Of course, triangulation fails absolutely when several intersection points fall within one beamwidth of the radar antennas. In this case, the targets must be resolved by identifying the range of the jammers.

To provide the resolution afforded by radar range data, an alternative means of measuring range is needed. Such range data may be obtained through observations made from two sites. The difference in the time of transmission of the jamming signals to the respective sites then gives the range. For a constant transmission-time difference, all possible sources lie on a hyperbola—neglecting errors due to altitude, which can be made small by choosing a suitable geometry. A mechanism for finding the time difference is (1) to send the signal from the remote site to the central site where processing is to be done and (2) to insert a time delay in the output from the central site of such a value that the signal after the delay is the same

3-16

as the signal sent from the remote site (except for a possible doppler shift). The difference between the inserted delay and the time to send the signal from the remote to the local site is the time difference corresponding to the given hyperbola.

To determine that the signal through the inserted delay is the same as the signal from the remote site, the two signals are multiplied together and their outputs are filtered. In most practical systems, the incoming signals are heterodyned to two different intermediate frequencies so that the output of the multiplier will be a signal at the difference of the intermediate fre-
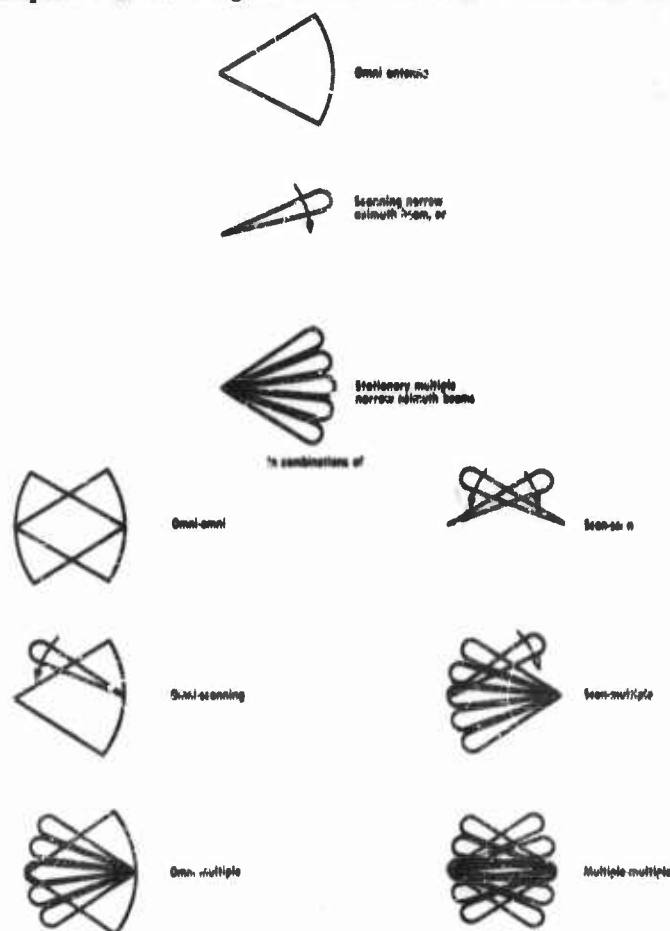
FIGURE 3-1.   Antenna configurations for passive correlation stations.

quencies, shifted slightly by whatever doppler difference occurs at the two sites. Several types of systems using this principle are shown in Figure 3-1. Some of the significant features of the illustrated configurations are as follows:

(1) *Omni-Omni* ($r - r$): This method requires two pairs of sites. Each receiver has all jammers as sources. Combining the two pairs to eliminate ghosts presents a computer problem.

(2) *Omni-Scanning Beam* ($r - \theta$): The available integration time is the same as in a radar with a similar beamwidth and rotation rate, and hence is longer than for the two-scanning-beam type. The penalties incurred are (1) that, since the omni has all jammers as sources, when the beam crosses several jammers the product of the input signal-to-noise ratio is low and (2) that a signal source emitting periodic components causes ghosts along the main beam. An additional problem is encountered in azimuthal selectivity because of the one-way use of the beam.

(3) *Rapidly Scanning Beam and Radar-Rate Scanning Beam* ($\theta - \theta$): This system suffers from having too short a time on target. Although the average signal-to-noise ratio during a scan is very good, false alarms occur when a small number of jammers are in each beam. If the rotation rate—hence, the data-output rate—could be slowed down, the method would be very good. Two pairs of sites would be able to provide velocity information, and this would allow for a reduction in data rate. The radars supply only position information, however, so the processing becomes more complicated.

(4) *Scanning Beam to Multibeam* ($\theta - n\theta$): The advantages of this type of system over the scanning-beam-to-omni system include the elimination of the ghost problem caused by periodic signals, and the decrease in the number of sources received at the site where the omni was replaced by the multibeam. The disadvantage is the increased complexity.

(5) *Multiple Beam to Multiple Beam* ($n\theta - m\theta$): The full multiple-beam-to-multiple-beam system includes all the advantages of the scanning-beam-to-multiple-beam systems and also allows for greater integration time. Again, the disadvantage is complexity.

Tracking radars provide much better angular resolution of targets by virtue of their narrow pencil beams. The simplest passive tracking system makes use of two radars spaced some 20 miles apart. One acts as the master radar and the other as a slave. The master radar locks onto and tracks a jamming target in angle only, since range data are not available. The angle data are transmitted to a computer, which causes the slave radar to sweep up the master radar beam. When the slave radar finds a jamming target, it stops and tracks it in angle. The video signal from the slave radar is transmitted to a correlation computer at the site of the master radar, where it is

cross-correlated with the video signal from the master radar to determine whether both radars are tracking the same target. If not, the search of the slave radar is resumed until the same target is tracked by both radars. Two versions of the system have been proposed, the so-called "narrowband" and "wideband" systems. In the former, correlation is used only to determine that the master and the slave radars are tracking the same target, and the range from the master radar to the jammer is computed from the angular position of the two radars and the known length of the base line between them (the so-called $\theta - \theta$ correlation system).

In the wideband system, the correlation time, τ, is used to determine the range of the jammer from the master radar as in the $(\theta - \tau)$ system. The terms "narrowband" and "wideband" refer to the bandwidth of the correlated video signals which must be transmitted from the slave station to the master station. In the narrowband case, only a 2.5 kc bandwidth is required so that audio communication channels could be used, while in the wideband system the normal radar-receiver video (several megacycles) bandwidth would be transmitted.

The narrowband system would provide range with an accuracy comparable to that of a normal radar tracking for ranges up to about twice the separation between the master and the slave stations in the cardiodlike area

FIGURE 3-2. Region in which range error is less than 20 yards (in yards from the master).

depicted in Figure 3-2. This would be true as long as the standard deviation of the angular tracking accuracies of the radars is of the order of 0.1 mil, which would be the case for a single jammer. Unfortunately, if more than one jammer were in the tracking radar beams, which are about 1° wide, the radar would angle-track the center of gravity of the collection of jamming signals. This would increase the errors in range determination by an order of magnitude, or more than would be provided by a

normal tracking radar in a nonjamming environment. The wideband system would overcome this degradation in range accuracy when a number of jammers are in the target-tracking radar's beams. As each jammer would have a different correlation time, this would establish its location on the corresponding hyperbolae with the two radar locations as foci. Thus, the wideband system would provide better accuracy of range determination and better resolution, as compared to the narrowband system.

In conclusion, it should be noted that the angle data furnished by a jammed radar might equally well be supplemented with range data from another radar operating at a different frequency. This radar might even be a simpler range-only slave device. The merits of such alternative solutions involve evaluation of enemy capabilities and relative equipment costs which are beyond the scope of ECM technology.

3.5 ECM and Ground Operations

3.5.1 General Functions

In the ground environment, ECM can be considered as another tactical weapon for a local commander's selection and employment. One might conceive of strategic uses of electronic countermeasures, such as widespread deception through transmission of false messages, but since these uses are limited in scope, only tactical applications of electronic countermeasures will be considered here. Tactical weapons should be employable down through the basic combat units. In a ground operation, each of these units, when operating as a combat unit, rarely occupy or are concerned with more than several square miles of ground. The ECM tactical weapon, because of its electromagnetic properties, can have effects on areas other than those occupied by the combat unit and its immediate opponents. It is, therefore, necessary that restrictions be placed on the free use of the ECM weapon by the local commander. The ECM functions most closely related to ground-warfare operations include: (1) variable-time-fuze detonation, (2) malfunction of enemy drones, (3) ECM of unenciphered tactical communication nets, (4) ECM of enemy mortar and artillery radar, (5) ECM of enemy electronic surveillance, (6) ECM of enemy tactical precision bombing, (7) antijamming of enemy repeater ECM, and (8) ECM of enemy missile guidance. The reader will note that several of these functions are also associated with environments other than the ground environment. However, the types of problems associated with these functions and the ground warfare environment will be commented upon briefly.

3.5.2 Variable-Time (VT) Fuze Detonation

A target can often be damaged to a maximum extent by detonating an artillery shell, bomb or missile at some height above the target area. There are several types of VT fuzes which can effect detonation of a warhead at a predetermined height; the only type of fuze considered here is the radio frequency type since other types are not discussed in this book.

The effectiveness of variable-time fuzes may be reduced first, by causing the fuze to predetonate at such a distance as to cause little or no damage to the intended target and second, by causing the fuze to "dud" and allow-

ing the warhead to detonate on impact. It is obviously desirable to predetonate a warhead at a range such that no damage is inflicted on the target. The second course, impact detonation, is less desirable, since in many cases the target still can be severely damaged.

Repeater jammers are the only effective countermeasure against the sophisticated, rail-frequency fuzes. Special problems associated with countering VT fuses include that of antenna interaction and that of providing sufficient power in a linear repeater (discussed in Chapter 16). In general, the area protected by a single jammer is limited, and one is forced to choose between very large equipments to protect a number of army units in large areas or a number of smaller equipments to protect units in small areas.

3.5.3 Malfunction of Enemy Drones

Serious consideration must be given to the use of electronic countermeasures against surveillance drone systems since they can prove to be real threats to forces being observed by such systems. Countermeasures might be successfully applied to drone guidance or navigation systems and to communication links between a drone and its control station. (Problems involved in the application of countermeasures to sensory systems carried by drones are discussed in Section 3.5.6.) Self-contained navigation systems may prove difficult to jam or to deceive, while navigation systems which rely on control stations may prove to be susceptible. Jamming of navigation or communication links to a drone may prove to be a profitable endeavor since a jammer transmitter can have a decided range advantage (and consequently a power advantage) over the navigation or communication transmitter. However, the range advantage can be offset through the use of directional antennas, sophisticated modulation techniques, or codes.

3.5.4 ECM against Unenciphered Tactical Communications Nets

The jamming of tactical communication transmission is a complex problem; the effectiveness of such jamming is also difficult to evaluate. Three factors bear upon the use of communications jamming. The first factor is that of jammer power. In general, high powers or height advantages are required because the jammer transmitter is often located at a greater distance from the target receiver than is the communications transmitter. The power requirements are increased still further by propagation and terrain-shielding problems (problems which arise because of the location of both receiver and jammer on the ground). A second factor is that of interference with friendly communications. Many of the friendly communication links will be operating in the same portions of the frequency domain as the

enemy communication links, and it would not be possible to eliminate completely interference with the friendly communication systems. Interference due to jamming can be reduced through the use of such devices or techniques as directional antennas and signal correlation schemes. Consideration must be given, of course, to the effect of these devices or techniques on such things as complexity, maintenance, weight, and volume. Still a third factor to be considered is that of intelligence, which requires the evaluation of the relative worth of monitoring enemy communication links against jamming of these links.

From technical and tactical standpoints, one has three possible choices for jamming tactical communication nets. The most straightforward procedure is to place high-powered jammers behind friendly lines and to transmit appropriate jamming signals into enemy territory where the target receivers are located. The high power required to produce effective jamming is a disadvantage, as is the increased interference with friendly communication nets caused by the high jammer power. A second possibility is to plant expendable jammers in enemy territory in the vicinity of the target receivers. Interference with friendly communications links is reduced because of increased distances from the jammers. A major problem is the development of satisfactory techniques for the sowing of expendable jammers in desired locations. A third possibility for jamming ground-based communication links is the use of airborne jammers. In addition to such tactical problems as radiating the jammer power to the desired areas, all the difficulties associated with operating and maintaining airborne electronic equipment must be overcome. An advantage possessed by airborne jammers is the decrease in terrain shielding effects normally encountered in ground-to-ground jamming.

3.8.5 ECM against Enemy Mortar and Artillery Tracking Radars

Technically, the problems faced in jamming mortar and artillery tracking radars are the same as those faced in jamming other types of tracking radars. However, a special problem is involved in that the jammer often incurs a power disadvantage because of the geometry of the jamming situation. When enemy mortar and artillery tracking radars are placed in defilade, it often becomes impossible to place jammers at vantage points; the radar antennas are sufficiently shielded by the intervening terrain so that only small amounts of jammer power are intercepted by these antennas. If the jammer power output is increased to overcome the losses, the jamming transmitter may be of sufficient nuisance to become a prime target for counter-countermeasures—possibly in the form of direct fire from artillery or a missile system.

Other countermeasures schemes can be considered for use against mortar and artillery tracking radars. For example, large quantities of chaff could be used to mask shell trajectories. The chaff could be dispersed by rounds fired previous to the beginning of prolonged bombardment. The problems involved would include determining the quantities of chaff required, the chaff fall rate, and the effect of weather (winds) on the chaff distribution.

3.5.6 ECM against Electronic Surveillance Devices

Electronic surveillance devices used in ground warfare may be either ground based or airborne. Ground-based electronic surveillance devices consist primarily of reconnaissance (intercept) receivers, MTI radars, and infrared scanners. The same problems attendant to countering tactical communication nets exist for countering ground-based intercept receivers. A jammer may not be able to range disadvantage with respect to a friendly transmitter being monitored by the enemy, but high jammer powers may be required to overcome propagation losses. The problem of interfering with one's own transmissions and the question whether jamming or monitoring the enemy transmissions is more valuable also exist. Ground-based MTI radars operate as line-of-sight devices, i.e., the radar antenna must point at the area under surveillance. Because these radars must essentially operate out in the open, they can more easily be subjected to countermeasures than can the intercept receivers. (Obviously, the location of a MTI radar can be more readily determined than the location of an intercept receiver). Difficulties will be encountered in jamming these radars because they can be made to operate at signal-to-noise ratios less than unity and because the jammer radiation will often be subjected to high propagation losses. These radars are made to operate only intermittently.

Confusion techniques also might be used against MTI radars. "Windmills," rotating reflectors, might be tailored to simulate the characteristic radar echoes caused by personnel or vehicles in motion.

Airborne surveillance devices include MTI radars, mapping radars, intercept receivers, and infrared scanners. The radars are usually area search devices and may be either side looking or forward looking. Radar performance can be degraded by jamming, but one may be forced to use a tracking jammer; otherwise, the degradation interval may be short. Intercept receivers are indeed susceptible to electronic countermeasures when these countermeasures are designed to take into account the applications and modes of operation of the receivers. Infrared scanners, being passive devices, are not susceptible to jamming in the usual sense; however, decoys might be used to deceive the scanners.

8.5.7 ECM against Tactical Bombing Radars

Several aspects of ECM against bombing radars are discussed in Section 3.4.2. While that discussion deals primarily with the use of ECM against radars carried by strategic bombers, much of the discussion is applicable to the problems of ECM use against tactical bombers. Obscuring real targets and providing false target information can be used in the tactical situation as well as in the strategic situation. In cases where only small areas need be defended and in which the enemy is using high-explosive bombs, a few low-powered jammers may suffice to provide a large measure of protection. There are problems, of course, of placing the jammers in desirable locations, of turning on the jammers at the proper times, and of maintaining equipment in the field.

8.5.8 Counter-Countermeasures against ECM Repeaters

ECM repeaters may be used against VT fuzes, radars, or communications equipment. The most straightforward antijam technique which can be used against ECM repeaters is saturation jamming, since most of these repeaters are essentially linear amplifiers. This technique consists of overloading the repeater by transmitting a large CW signal in the passband of the repeater; the jamming signal normally transmitted by the repeater is reduced. When operating ground-based repeater jammers, one must face, as in operating much other ground-based radio equipment, the limitations imposed by ground-wave propagation conditions.

8.5.9 ECM against Missile Systems

Tactical missile systems used in ground warfare can be placed in the categories of surface-to-surface, surface-to-air, and air-to-surface systems. Of these, the surface-to-surface systems are probably, as a group, the least susceptible to countermeasures. Many missiles in this group are of the ballistic type and are unguided or guided only during the early portions of their flights. Countermeasures might be applied should radio or radar controls be used from a ground complex during the guidance phase. Difficulties abound in applying countermeasures: the guidance complex may not be easily found; the range from a jammer to the guidance complex may be long; transmission times of the guidance radios or radars may be short; and security codes may be used in command links. These problems may be alleviated somewhat should a missile be radio or radar guided during its midcourse phase or terminal phase. The airborne missile may be more easily located than a ground-control complex, and the jammer-to-receiver range may be shortened.

Short-range, antiaircraft missile systems will be used on the battlefield

3-24 ELECTRCNIC COUNTERMEASURES

against aircraft, drones, and missiles. These systems will generally have, as portions of their guidance systems, acquisition radars and tracking radars. These radars can be subjected to countermeasures, but care must be exercised when these measures are applied. For example, an enemy could be alerted by an acquisition radar to the presence of a low-flying aircraft whose jammer is on; deeper penetration might be achieved with the jammer off. As another example, some ground-based tracking radars and some missile-borne seekers are capable of tracking noise; a countermeasure other than a noise jammer would have to be used against such radars.

Countermeasures may also be difficult to apply against tactical air-to-surface missile systems since many of these systems will utilize unguided or inertially guided missiles. The means for acquiring targets for these systems may be more susceptible to countermeasures than the systems themselves. An obvious countermeasure against electromagnetic-seeking air-to-surface missiles is to turn off the source of radiation being attacked. If this is not feasible, confusion techniques such as false targets might be employed.

3.6 ECM and Underwater Operations

Both submarines and surface ships depend highly on acoustic phenomena to help find their enemies and to deploy their weapons. Sonars (sound ranging) are used for intercept, identification, and tracking purposes; and sound receivers are used in acoustical homing devices. These devices and the countermeasure devices which can be used against them are described in Chapter 23. Radars and magnetic anomaly detectors are other devices used by surface ships and submarines to detect each other; these devices will not be described.

3.6.1 ECM against Surface Ship Weapons

Submarines can use countermeasures to advantage when attacking or when under attack by surface ships. Mounted or towed jammers can be used to reduce the effective range of a ship's sonars. Expendable jammers can be used as a screen behind which the submarine could maneuver while attacking a ship or when escaping from a ship's attack. Decoys might also be used to advantage during disengagement from battle by misleading a ship as to the submarine's position. Another use of decoys is to deceive the enemy into thinking that there are more submarines attacking the ship than there actually are. The more sophisticated decoys should be able to simulate doppler echoes, propulsion noise, wakes, and maneuvers. Typical problems when using or devising such decoys include those of programming a decoy's course and of simulating properly the varying echoes as a submarine presents its beam, its bow, or its stern to a ship's sonar.

3.6.2 ECM against Submarines and Torpedoes

A submarine can often detect with its sonic devices the presence of a surface ship before the ship can detect the presence of the submarine. One way to reduce this range advantage is to raise the noise level by means of towed noise makers or jammers. Some means of look-through must be provided for the ship's sonar devices if the stratagem of raising the background noise level is adopted. Expendable jammers can be used for screening purposes by a ship as well as by a submarine. Decoys can also be fired between submarine and ship, either to confuse those aboard the submarine or to become targets for homing torpedoes.

3.7 ECM in Space

The discussions so far in this chapter have dealt with the use of electronic countermeasures in war environments which have been explored and studied extensively. The expansion of military operations into space will introduce new problems into the operating perspective of electronic countermeasures. In general, the principles underlying the uses of electronic devices and electronic countermeasures techniques in space will be quite similar to the principles of their use on earth. This is true because the devices to be countered, whether on earth or in space, all fulfill in a broad sense some function of surveillance or communications. The discussion here of electronic countermeasures in space includes not only interplanetary space operations but long-range ballistic-missile and earth-satellite operations.

One way to examine the problem of countermeasures in space is to consider countermeasures in terms of the three functions of surveillance, jamming, and deception. Further, these functions can be applied from the earth or from space to vehicles or objects on earth or in space. Thus, the environmental combinations include space-to-space, space-to-earth, and earth-to-space. In all, the three countermeasure functions times the three environmental combinations result in nine possible countermeasure environments. Some of the space applications of countermeasures are: (1) electronic surveillance of objects on earth or in space; (2) jamming and deception of ground radars; (3) jamming and deception of ground communications; (4) jamming and deception of satellite communications (communications between satellites or between satellites and the ground); (5) decoys and penetration aids for launch, midcourse, and terminal phases of ballistic-missile flights; and (6) jamming and deception of AICBM complexes. In the last category, AICBM complexes may include such things as reconnaissance devices on satellites, satellite interceptors, infrared homers, early warning radars, fence-type detectors, and trackers. How rapidly progress is made in the development of countermeasures in these areas is somewhat

dependent upon the development rate of the systems to be countered. For example, the use of countermeasures by and against ballistic missiles has been studied extensively, while much less effort has been expended on the study of the use of countermeasures against satellite interceptors. In time, if and when satellite interceptors become a threat, more effort would be expended on the development of countermeasures for use against these interceptors. Problem areas, typical of the types to be encountered in the application of countermeasures in space are enumerated in the following sections.

3.7.1 Countermeasures against Electronic Surveillance of Objects on Earth

Sensors which may be used on surveillance satellites are photographic and television cameras, radars, infrared detectors, and radio receivers. Satellites will also have a means of communicating with some control center. A data link in the form of radio transmission might be used, or a capsule might be ejected. Both the sensors and the data links of satellite surveillance systems can be subjected to countermeasures. Whether the sensors or the data links or both are subjected to countermeasures depends on the nature of the particular reconnaissance system.

Jamming might be successfully applied against radio receivers and radars by utilizing tracking jammers. Some of the problems involved are determining the frequency band of the receivers or radars, generating and transmitting sufficient jamming power, and appropriately placing jammers so as to obtain coverage over the areas to be protected. Infrared surveillance devices would be difficult to jam since they are usually scanning devices. It would be impossible to maintain a jamming source within a scanner field of view for any length of time. Intense light sources such as lasers could produce halos on photographs, but the pictures would be obscured only over small areas. Deception techniques in the form of decoys or shields offer more chances of success as countermeasures against photographic devices.

Data links will be difficult to disrupt because, in most instances, radios operable only on command will be used. If these radios are operated only over the satellite user's territory, it will be difficult to place jamming or deception devices in positions where they can be used. If the nature of the command code can be determined, it might be possible to preempt the command link over friendly territory and to fill a satellite's memory devices with useless data. Again, what countermeasure technique is to be applied depends upon the system to be countered.

3.7.2 Countermeasures against AICBM Complexes

A point defense, ground-to-air missile system (as exemplified by Nike-

Zeus) and missile interceptors launched from a satellite are two examples of AICBM systems. There has been considerable study devoted to the development of concepts and techniques for decoy use against tracking radars of point-defense missile system. The simplest technique which has been considered is to fragmentize the final-stage fuel tank of a missile so that the warhead could not readily be distinguished from the fragments. The potential use of this technique and others has aroused sufficient concern so that counter-countermeasure techniques are being investigated. One of these techniques involve examining in detail the radar signature of each object to determine differences in the spectrums of the radar echoes. Another technique involves examining the trajectory profiles of each object to determine if their mass-size ratio is like that of a warhead. As the counter-countermeasure techniques are improved, old countermeasures must be refined and new ones must be devised. A new technique, for example, might involve the use of an expendable jammer. The jammer may have to be expendable to overcome track-on-jamming capabilities of a radar.

Several satellite weapon systems are being considered for use against intercontinental ballistic missiles during the missiles' boost phases. In these systems, infrared-seeking missiles are launched from a satellite at the ballistic missiles. Decoys immediately suggest themselves as a countermeasure to both an infrared search device and to the infrared-seeking missiles. How many decoys should be fired, when they should be fired with respect to the firing of the missile, the decoys' infrared radiation characteristics, and their effectiveness and cost must all be determined. Because the infrared-seeking missiles use tracking devices, the possibility of developing and employing a jamming device exists. Whether or not such a device could be used would depend on the infrared seeker's characteristics.

Part II

Signal Intercept

4

Operational Objectives of
Intercept Systems

L. A. deROSA, E. FUBINI, J. VOGELMAN

4.1 Purpose of Reconnaissance

Intercept systems are one form of reconnaissance and possess many of the characteristics of reconnaissance systems in general. Reconnaissance is a collection of information on the facilities, capabilities, and intentions of a potential or actual enemy. It is the mission of reconnaissance to measure the effectiveness of these facilities; to estimate their reliability; to determine deployment and changes in the enemy's strategy and tactics. Effective reconnaissance leads to effective redeployment and modification of one's own strategies and tactics.

It is not enough, for instance, to determine the existence of a particular enemy facility. Data must be collected regarding its operational employment, usefulness, location, frequencies, power, field of coverage, rate of transmitted information, over-all reliability; and its immunity from jamming, detection, crypto-analysis, and natural and artificial interference. In general, its threats and its points of vulnerability must be determined. These factors play a fundamental part in evaluating the importance of the facility to the enemy, and in formulating one's own plans.

It is impossible to properly evaluate the operation of ferret intercept systems without remembering that the collection of information regarding an enemy's electronic and communication facilities is only one of the many intelligence missions that must be treated simultaneously to gain knowledge of enemy strategy and facilities, and changes in either or both. Total intelligence results from the following types of collections.

A. Visual
 1. Photo
 2. Television
 3. Human vision
B. Communications interception
 1. Traffic analysis and analysis of uncoded messages
 2. Cryptographic analysis
C. Electronic Intelligence
 1. Deployment
 2. Technical intelligence
 3. Tracking of electronic sources
D. Radar Intelligence
 1. Mapping
 2. Search and tracking
E. Infrared detection
 1. Image forming
 2. Search and tracking
F. Collateral information

Any one of these types of intelligence, though powerful in itself, may have a very much increased importance if used in connection with another. One discovery by any sensor can trigger a collection procedure by one of the other sensors, a procedure designed to confirm or discount the conclusions that may be drawn from the initial collection.

Thus, one type of sensor will give rough information which can act as an alarm by which detailed information-gathering devices can be turned on. The infrared detection of a missile, for instance, will alert the sensor whose mission it is to detect guidance signals. Another example: the receipt and analysis of an electromagnetic signal can establish the reliability and accuracy of a covert source.

Electronics reconnaissance has immediate use in dictating a reaction capacity; it has a long-term use in the strategic evaluation of the enemy's capabilities. The value of the data acquired from this type of reconnaissance depends ultimately, of course, upon their use. The following pages will discuss problems of collecting ELINT information, and some of the knowledge to be gained from such information.

4.2 Distinction Between Reconnaissance and Communication

Electronics reconnaissance differs significantly from communications reconnaissance. For one thing, the amount of information required for successful electronic intelligence (ELINT) is much less than that required for successful reconnaissance of conventional message-carrying channels.

Effective electronics reconnaissance requires the surveillance of very large frequency bands, yet such surveillance is within the range of possibilities.

There are facets of electronics reconnaissance that are not found in communications reconnaissance.

Note, for instance, that the geographic location at which a signal is intercepted, its direction, its rate of change of bearing, etc., represent data that are not usually relevant to conventional communications reconnaissance. Such data are relevant to navigation systems; but there the data are put to a different use from the use to which they are put in electronics reconnaissance.

The reaction of an enemy to a particular penetration attempt conveys useful intelligence information—it can tell much about his capabilities and his plans for putting these capabilities to use. In conclusion: *electronics reconnaissance is not the same as message reception*. This difference has been neglected in many recent reports.

4.3 Relation Between Data Required and Their Intended Use

It is perhaps helpful, but often forgotten, that the fundamental usefulness of reconnaissance is a function of the use to which the data are put. *In general, mere recording of data postpones—not serves—the analysis purpose.* The recording of data that are not analyzed is often a greater waste than merely the obvious waste of recording capabilities: useful information may have been neglected during the recording process. A useful rule might be that information must be recorded only if it is clear how the recording will be used.

The processing and analysis of electronics reconnaissance data are often so time-consuming that is it only the availability of computing machines that makes possible efficient reconnaissance capability. Unfortunately, the types of analyses required vary a great deal and depend upon the type of electronic facility being observed. Some analyses require manual and direct observation. Others require coordination of a number of sensors. This is a fundamental part of the electronics reconnaissance problem. For example, to establish by analysis the existence of a beacon responder system, it is necessary to obtain reconnaissance intercept on both the interrogator and the responder. Because of the character of electronics reconnaissance, it is always necessary to work from a large mass of often useless data in order to find the few items likely to have immediate importance and direct usefulness. Separating, sifting, and sorting are some of the most difficult and important tasks in analyzing electronics reconnaissance data. For example, repetition of a detected signal, which will occur many times, only confirms its existence; it does not add new data.

It is not a purpose of this chapter to explore the final destiny of reconnaissance data after completion of the analysis process; but one should remember that dissemination of intelligence data is just as important a part of the over-all system as the collection and analysis for those data.

4.4 Operational Use of Reconnaissance Information

The detection of electronic facilities installed by an enemy presents one of the best methods of reaching conclusions regarding his deployments of weapons, aircraft, missiles, and other tools of air, sea, and ground battle. In the case of the Strategic Air Command (SAC) for example, knowledge of the deployment of weapons permits an optimum planning of penetration for a maximum chance of reaching the target safely; in the case of ground forces, it permits evaluation of the direction of enemy effort and suggests tactics for reaching weak spots in enemy deployment. The proper balancing of electronic countermeasures and active weapons, and the over-all conduct of the battle can be very greatly implemented by detailed knowledge of the employment and deployment of the enemy's electronic capabilities. Perhaps one of the most striking examples of the use of electronic reconnaissance to determine enemy reaction is that involved in aerial warfare.

Electronic countermeasures can be employed against electronic weapons with a substantial power advantage. In the early days of electronics, the number of electronic weapons was small and their sophistication was not that of today's electronic weapons; consequently countermeasures could be effectively employed with relatively small powers. For instance, powers of the order of magnitude of a tenth of a watt per megacycle were sufficient to effectively jam the German radars during World War II. Besides, the German radars occupied an over-all bandwidth of no more than 500 megacycles; for that reason, a total of 50 radiated watts could do a satisfactory job against the German defenses in 1943 and 1944.

The situation has been changing rapidly. With the progress of electronics, the number of electronic weapons has increased immeasurably in the last ten years; also, the sophistication of these weapons has made jamming steadily more difficult. At this writing (1959), powers of the order of 10 to 20 watts per megacycle are required to effectively interfere with the operations of some of the modern types of radars. The availability of new tubes and components has permitted the use of wider and wider frequency ranges so that today as many as 10,000 megacycles are available to an electronic-minded enemy. One could, therefore, say that a brute force jamming effort, based on no information whatever of the enemy's intentions, technical capabilities, and deployment, would require 200,000 radiated watts. That means approximately 1,000,000 input watts or 1000 kilowatts. Input powers of this order of magnitude are not available to present-

day aircraft; thus, it is imperative that accurate information be made available to operations planners for effective use and deployment of chaff, deception, and jamming capabilities.

If we knew the exact frequency ranges in which an enemy would use his capabilities, we could reduce by a factor of about 50, and perhaps 100, the electronic countermeasure equipment needed for an aerial strike against his country. Although it is not the only method of obtaining such intelligence, electronic reconnaissance is one important method.

The sophistication of radar and electronic devices is such that some of their important characters cannot be revealed by passive listening; however, the information gathered by passive listening contains the majority of important data. For example, the availability of coherent or noncoherent "Moving Target Indication" cannot be determined by passive intercept of signals, though the probability of its existence may be deduced from the stability of the transmission. The data that passive listening gather, coupled with the collateral information, might provide knowledge far more difficult to acquire than the knowledge passive listening might be expected to provide.

Another example of operational use of electronics reconnaissance can be described in terms of immediate reaction capabilities. It is well known that electronic decoys are particularly useful in the presence of passive locators; decoys can prevent an enemy from effectively triangulating on our own jamming or repeater transmitters. On the other hand, the number of useful decoys that an aircraft can carry during a deep penetration mission is limited. It is, therefore, important that a preprogrammed set of sensing devices be supplied to an aircraft to permit the release of decoys at times of maximum usefulness.

Operational use of ELINT can be demonstrated by examples of other types of warfare—for instance, in the protection and detection of submarines. Recent maneuvers have proved that a submarine can, by passive listening to r-f radiations, effectively detect the presence of enemy aircraft and, in many cases, of convoys or enemy forces. For this reason, passive reconnaissance in naval operations is of utmost importance.

The likelihood of limited warfare is becoming greater. The theaters of limited warfare will probably be located in regions where the density of electronic transmissions is normally low. Because of the low traffic density in those regions, the problem of interception would be greatly eased. It is highly likely that increased electronic traffic would mean increased enemy troop concentration. It can be expected that the study of electronic and communication intelligence would play as important a part in enemy capabilities and in guiding our reaction during a limited warfare ground battle as it would in determining the optimum use of electronic countermeasures in a possible strategic air operation. This will be particularly true and important when build-up

of electronic weapons can be used to measure the help that a highly technical enemy is giving to an underdeveloped country.

The continuous increase in the use of electronics for communications, navigation, and other military purposes in all countries of the world has brought about a corresponding increase in the size and importance of electronic warfare. As a type of war, electronic warfare takes on many of the characteristics of conventional warfare with conventional weapons. Here, also, the offense and the defense have the upper hand at different times; and the knowledge of the enemy's capabilities is essential for planning our own reactions and strategies. Any general statement made today regarding the relative superiority of particular countermeasures, or of particular electronic weapons against countermeasures, is likely to be proved wrong a few months after it is made. *The result of a particular test of a particular countermeasure against a particular set of weapons can never be considered as over-all absolute proof of relative importance or usefulness,* nor can it be extrapolated to future conditions without severe limitations.

4.5 Electronic Maps

A map of the electromagnetic signals that can be heard today by an aircraft flying over the United States would include a wide variety of transmitting sources. One could find landing aids, airport radars, television stations, radio broadcast stations, Air Force, Army, and Navy radars and communications, commercial radars, airborne radars, and navigtion aids of all kinds. By and large the density of its electronic signals is a measure of the technical development, the population, and the industrialisations of an area. In general, the same type of statement can be applied to a battlefield area. This continuously increasing use of electronics has increased the usefulness of electronic reconnaissance and intelligence to all types of warfare, so has it increased the difficulties of analysis.

There are some not-so-obvious uses of data obtainable from electronics reconnaissance. The analysis of an enemy's deployment of electronic devices can often be used to give a direct measure of his production capability. By means of the observation of the time interval between the appearance of the first of a new type of equipment and the appearance of a series of such units, it is possible to estimate his ability to muster industrial production. If, by ferret reconnaissance or by other means, it is possible to obtain information of a particular technical development in its early stages, the time interval between the early development stage and the installation of the first operational equipment gives invaluable information regarding the lead time that the enemy requires to develop and make operational use of equipment.

In analyzing different complexes—missile ranging sites, atomic plants—
the use of electronics reconnaissance has become of greater and greater
importance. The observation of our own missile ranges has disclosed a close
correlation between signal traffic and missile launchings; the analysis of
these signals, their types, and their schedules conveys to a skilled analyst
important information regarding missile operations (Figure 4-1). Infor-
mation regarding missile capabilities can be effectively obtained by radar
observation of missile trajectories, especially when correlated with the
interception of telemetered signals. In the aggregate it can be stated that
the continuous increase in development of electronic facilities has multiplied
the importance of their interception and has increased by very large factors
the amount of information which can be drawn from it. Electronics and
communications reconnaissance may contribute as much or more useful
data than photographic reconnaissance of the same area. Actually the prob-
lem of photographic reconnaissance can be greatly eased by the use of
electronic devices: the careful photointerpretation of large areas of land is
not easy unless data as to where the observer should concentrate his atten-
tion are available. Electronic emitters are, by their very nature, almost
impossible to camouflage and for this reason their locations are often
indicative of the presence of targets of importance for photographic recon-
naissance. It is conceded that targets exist that do not emit or reflect
electromagnetic signals; therefore, this mutual support between sensors is
not of universal use. Experience and analysis indicate, however, that there
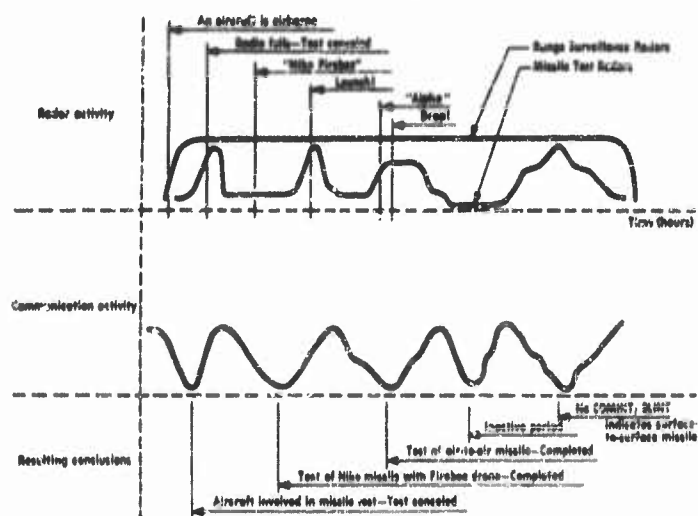
FIGURE 4-1—Typical Missile Test at Cape Canaveral

are a number of cases where reconnaissance can be increased in effectiveness by combinations of all types of electronic radar and photographic sensors.

4.6 Operational Exploitation of Radar Weaknesses by Means of Electronics Reconnaissance

As stated in the previous section, one of the fundamental properties of electronic signals is that they cannot always be camouflaged, and a weakness of radars is that their signals must be on the air in order for them to be effective. An alert enemy with good means of reconnaissance can effectively employ this weakness to his advantage. The best way of explaining how this can be done is to take as an example an Early Warning System. This system might use very powerful radars that can be intercepted at a long distance. If they were turned off for maintenance of malfunction, the enemy could easily be alerted by his reconnaissance and choose that time as the best one to launch an attack. Furthermore, the enemy could from time to time probe the system by using jamming or decoying to determine our reactions to such signals. He could, for instance, determine the range over which we can tune a transmitter in a given time. This information would be invaluable to an enemy planning attacks against our defenses. In a similar way, naval task forces and aircraft flights could determine the alertness and operational procedures employed by an enemy to protect his shore lines. Continuous but intermittent jamming or deception operations by an alert enemy might be effective as a countermeasure of Early Warning Systems if, by forcing the defenses to "cry wolf" enough times, they can reduce the national confidence in such a weapon.

4.7 Communications Intercept and Traffic Analysis

It is well known from newspaper reports and historical records that the interception of enemy communications has been of vital importance many times in determining the outcome of a battle or of a war. It is not the purpose of this chapter to discuss the methods, the procedures, or the collection devices used to intercept communications; nor is the intent of this chapter to discuss the deciphering of enemy messages. Some aspects of communication reconnaissance are, however, closely enough related to electronics reconnaissance to warrant separate consideration in this chapter.

The value of traffic analysis can be demonstrated best by assuming that the coding methods used by a hypothetical enemy are such that decoding is impossible. Despite the impossibility of decoding messages, communications intercepts are very important. Many parts of the message other than the content convey information. The type of transmission; addressees; the number of messages; signatures; length of the messages; the types of transmitters employed, their powers, ranges, types of modulations, the type of code, its apparent complication and sophistication—all these are charac-

eristics that can permit sorting and classification in a way similar to that employed in the analysis of conventional ELINT. The relative frequency of multiple-address messages to a group of users will establish their common interest. Then, later-acquired knowledge about one of the group might indicate the interests and missions of the others in the group. A sudden increase of traffic between a logistic and an operating unit may give advanced warning of an important operation. To avoid warning the Germans of the impending invasion of Normandy during World War II, many dummy messages were sent starting several months ahead of D-Day to establish a traffic level identical to that required just before and on D-Day.

4.8 Alarms and Analysis

One of the fundamental purposes of reconnaissance is, of course, to recognize the existence of unexpected, new, or particularly important signals in a large mass of traffic. It is important to understand that electronics reconnaissance has two separate and distinct missions. The first is that of determining that a particular signal does not belong to any known class. The second is that of analyzing the character of signals to determine in detail all their characteristics.

The distinction between merely discovering the existence of signals and analyzing them is an important one. It is important in the operation of ELINT programs, and it is important in the designing of equipment aimed at the detection of ELINT. It is much easier to make and design a number of devices capable of alerting an operator when something either expected or unexpected occurs than it is to make devices capable of making analyses of all kinds of signals. However, if the distinction between detection and analysis is kept in mind, it is possible not only to process large amounts of electronic traffic, but also to design devices that select out of this traffic the very small percentage that is unfamiliar or for other reason deserving of special notice. An alarm may then be sounded, a strike reconnaissance mission initiated, a missile or decoy launched, or finally—most important— from the intelligence point of view a number of analysis devices may be turned on with or without recording capabilities.

It is only by this analytical means that effective use can be made of wideband recording, or of human observation of oscilloscope patterns. In the presence of the large number of electronic emitters with which a civilized country deals today, and which an aircraft or a ground-based station is likely to encounter in a theater of operations, analytical procedures are necessary if electronics reconnaissance is to be effective. Present intelligence estimates lead us to expect the intercept of at least 4000 important signals every second by aircraft flying over a heavily defended area at 50,000 feet (Figures 4-2a and b).
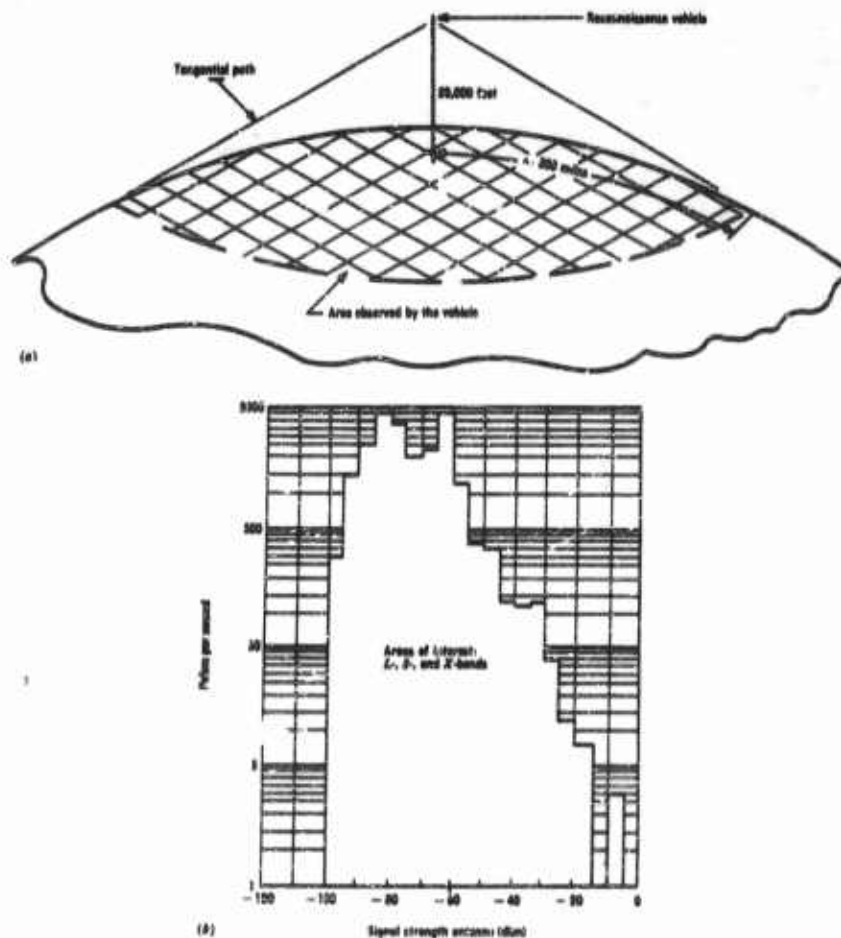
FIGURE 4-2—Signal Densities in Representative Areas

4.9 Conclusion

Electronics reconnaissance and its operational employment are increasing in importance because of the continuous increase in the number of radars and other electronic emitters, and because of the increasing sophistication of these weapons. The function of electronics reconnaissance devices has been extended from that of simple aircraft detection to the detection of warlike intentions on the parts of potential enemies, to battlefield surveillance, to the determination of enemy missile launchings, and to the study of enemy production and industrial capabilities, political moves, and covert data collection.

5

Signal Environment Study

A. E. HALTEMAN

5.1 Introduction

Operational environment plays a large role in the design of intercept equipment for the detection, location, and recognition of a signal associated with a particular piece of radiating electronic equipment. The environment of primary concern is that often referred to as the electronic or signal environment. When a radar, a guided missile, a bomb fuse, or an electronic countermeasure system must operate in an area in which there are many electromagnetic radiators it is necessary to know the nature and amount of the interference that may be present in order to design equipment that will function properly. A variety of studies aim at describing or measuring in some manner the expected pulses or sequences of pulses from the environment which could actually cause malfunction of the receiver or its associated equipment. The material in Section 5.2 describes briefly three such studies and compares some resulting estimates of the signal density.

In Section 5.3 a procedure for obtaining tactical electronic intelligence (ELINT) information from a complex signal environment is described. The device recommended for use is a simple pulse counter.

5.1.1 A Brief Summary of Recent Work

The first real awareness of an interference problem occurred during World War II. A natural outgrowth was a military request for the detailed description of what was called the "Radar Order of Battle" for two opposing field

armies. These studies were done with the cooperation of consulting field army personnel and took the form of detailed charts listing all radiating equipment associated with a field army and geographic maps showing typical deployments of the tactical units with their associated radiating equipment. As new equipment was developed and different army organizations were planned a projected radar order of battle for the 1960-1965 period was obtained. A Rand report (Reference 1) was one of the first of these.

The most complete list of radiating equipment associated with the U.S. field army and the opposing enemy field army now available is Volume II of the *Final Report, Project Monmouth I* (Reference 2). The geographic deployment of the electronic equipment is that which would be expected in the strategic European area in Germany. The implied large number of signals that would be present and recommendations for reducing the possible interference are presented for both the communication and the radar frequency bands.

Project Monmouth I and other studies point out clearly the possibility of unintentional interference that may continue to occur in the future and it describes the variety of signals which may need to be analysed by intercept equipment. These studies fail to answer the question often asked by the design engineer: how many signals having similar parameters will the system be required to handle simultaneously? More recent efforts may be able to offer an answer to this question.

Myers and Van Every (Reference 3) developed a method for predicting the signal density to be expected to arise from a given distribution of radars. Signal density is defined as the total number of pulses received above the receiver's sensitivity level. The main body of the report is devoted to predictions of the signal density to be expected in L-, S-, and X-bands during an active wartime situation in Europe in the 1960 period.

In a report done at Melpar, Inc. (Reference 4), an attempt has been made to present the electromagnetic environment within which tactical countermeasures must operate. The aim of the report is to establish realistic technical and operational requirements for such countermeasures.

The approach involved setting up a geographical model of a typical army area and surroundings with a "realistic" defense deployment of men and field equipment. A list of all devices whose radiation contributes appreciably to the electromagnetic environment along with the nature and volume of traffic handled by these devices during a typical battle was compiled. Using a digital computer a determination of the signal complex as seen by one or more intercept stations strategically located within the field army area was made. The computer was programmed to include certain effects from sky waves, from terrain reflections and from other propagation phenomena. An

experimental program held in Arizona offered some empirical inputs to the over-all program. This is a very comprehensive study but unfortunately is time and location limited. It should however be very useful in the design of equipment for use in the immediate future.

5.2 A Description of the Three Signal Density Studies Done at the Sylvania Electronic Defense Laboratory

Signal environment studies at Sylvania were instituted with the primary aim of obtaining a method for estimating the number of signals a receiver-analyzer system will be expected to handle simultaneously. Such information is essential for the design and development of such equipment. In addition the effect of the various receiver and environmental parameters on this number is of interest.

5.2.1 A Monte Carlo Technique for Determining the Signal Density in a Tactical Situation*

The engineer designing intercept systems (composed of an antenna, a receiver and a pulse analyzer) has at his command some system parameters which if judiciously chosen will reduce the number of signals the analyzer must handle. Those chosen as most important for this study are: antenna beamwidth (assuming a cosecant square antenna pattern), antenna gain, receiver sensitivity and receiver bandwidth. These parameters limit the signal density which here is taken to mean the number of independent signals that appear at the input terminals of the signal analysis equipment in a specified interval of time. Signal environment for this study is defined as the total ensemble of signals in the electromagnetic spectrum that the intercept equipment is capable of receiving.

Assumptions and Methods Used to Build the Monte Carlo Model. To determine the signal density of the output of a receiver the signal environment described by Hiebert of Rand Corporation in R-280 was used. The study was limited to the 373 various types of S-band equipment tactically deployed along a theoretical battle front chosen as the east-west German boundary between Witzenhauser at the north and Coburg at the south. Opposing the U. S. field army is a USSR field army with estimates of the kinds and locations of their electronic equipment. Within these opposing armies exact location of the various radiators had been pinpointed.

In the U. S. Army all equipment with tunable magnetrons was arbitrarily assigned permissible subbands in which to operate. For a particular radiator, the subband in which it would operate was determined by random

*See Reference 5.

selection. In the case of fixed magnetron equipment it was assumed that nominal dimensions of the cavity are specified on the basis of a design frequency centered in the band of the magnetron. A gaussian distribution was assumed to describe the exact operating frequencies of a given magnetron type. The operating frequency for a particular radiator was then selected from the distribution.

For the opposing forces all S-band equipment types were assumed to be a Soviet equivalent of the AN/MPQ-10. The method described above was used to assign frequencies except for slight modifications needed to account for the tighter tolerances common in Russian manufacturing.

A gaussian curve was fitted to the half-power beamwidth of the antennas to the isotropic level. Isotropic level was assumed outside the main beam since the primary object was to determine the probable transmitter antenna gain in the direction of the intercept site. For tracking radars this pattern was assumed in both elevation and azimuth. However this had no effect on the signal density since such radars are more likely to be pointed upward. The pointing direction of the various search radars relative to the intercept site was made by assigning an equal probability to each direction and sampling randomly from the uniform distribution.

The intercept site was selected at a point 30 kilometers behind the main line of resistance near the center of the army area and on reasonably high ground served by a road.

Assuming the transmitter antenna and receiver antenna each at a height of 18 feet above the ground, transmission losses were computed using the radio range equation for line-of-sight transmission and a modification of it for transmission beyond the radio horizon. The contributions to the signal environment at the intercept site for any given radiator were obtained and plotted on a graph showing frequency versus free space power at fixed azimuth.

An intercept receiver can be thought of as a narrow-pass filter whose input, in this case, is the signal environment and whose output is the signal density. Such a filter is dependent on the receiver antenna pattern, sensitivity, and bandwidth. By choosing reasonable values of these parameters it was possible to make a cutaway template on the same scale as that used on the receiver input charts. Placing it on the frequency versus free space power chart an antenna scanning in frequency was simulated. The signal density was obtained by counting the number of points that appear in the antenna pattern template at each frequency of interest and in azimuth increments of 4 degrees over a 90 degree sector. (See Figure 5-1.) Figure 5-2 is a sample chart of the resulting signal densities.

The signal density figures in Figure 5-2 show the number of different
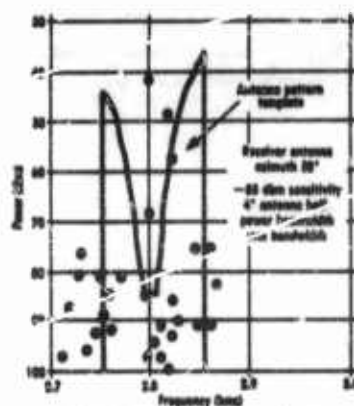
Figure 5-1 Receiver Input Environment

signals that may appear within the band-width of the receiver during the interval of time that the receiver is "looking" at the designated azimuth and frequency. The sample taken from the study is from the region of maximum signal densities, right being the highest value obtained in the study. Since no enemy equipment appears in these frequencies in the Rand Model all the signals received are friendly equipment.

It is recognized that many simplifying assumptions were made, but it is felt that the approach is reasonable and the results have the right order of magnitude.

Receiver Frequency Band mcs	Receiver Antenna (azimuth degrees)														
	32	36	38	40	42	44	46	48	50	52	54	56	58	60
2785	2	2	2	2	3	2	2	2	2	2	3	2	3	2
2790	5	5	5	5	5	5	6	5	6	7	6	6	5	5
2795	4	5	4	4	4	4	4	5	6	6	6	6	5	4
3800	3	3	2	2	2	2	4	6	8	5	4	4	3	3
3805	2	2	1	1	1	1	1	3	3	3	3	2	1	2

Figure 5-2 Signal Density Sample Chart

5.2.2 A Deterministic Model for Estimating Signal Density*

The present signal interference problems in the crowded communication bands and the rapid advance of tube technology have set in motion a trend toward controlled use of the military frequency bands. Thus a reasonable assumption from which to start a signal density study is that the probability of finding a transmitter of a given type in any square meter of the relevant part of the earth's surface is the same as in any other square meter. This assumption, though not completely realistic, has the advantage of no longer

*See Reference 6.

requiring the precise type, location, and frequency of each of a long list of transmitters. Also it avoids a series of rather tedious computations which account for the effect of each transmitter on the signal environment at the intercept site.

Assumptions and Mathematical Model Used. The power p_r delivered to input terminals of a receiver which is at distance r from a transmitter of power P_i on a direct line of sight is given by

$$p_r = \lambda^2 G_r G_i P_i / 16\pi^2 r^2 \tag{5-1}$$

where λ is the wavelength of the signals, G_r the gain of the receiving antenna and G_i the gain of the transmitting antenna. If we set p_r equal to P_r, the threshold sensitivity of the receiver, we may solve Eq. (5-1) for the maximum range of the receiver within the line of sight, obtaining

$$r = (\lambda/4\pi) \ (G_r G_i P_i / P_r)^{\frac{1}{2}} \tag{5-2}$$

Because the gain G_r is a function of the azimuth angle ϕ for a given receiving antenna orientation the range is also a function of ϕ. The area under the curve $r(\phi)$, plotted in polar coordinates, defines the region within which all transmitters of effective power $G_i P_i$ are picked up at the receiver. The area enclosed by this curve is

$$A_i = (\lambda^2 G_i P_i / 32\pi^2) P_r \int_0^{2\pi} G_r \phi \, d\phi \tag{5-3}$$

Let n_i be the number of transmitters of effective power $G_i P_i$ per unit area whose carrier frequencies fall within a band of width B. If $b(b \leqq B)$ is the bandwidth of the receiver the expected density of those transmitters whose carrier frequency falls within the passband of the receiver is bn_i/B and the expected number of such transmitters within area A_i is $bn_i A_i/B$. Summing over all types of transmitters we find the expected signal density is

$$N = \sum_i bn_i \frac{A_i}{B} = \frac{\lambda^2 s b}{16\pi P_r B} \cdot \frac{1}{2\pi} \int_0^{2\pi} G_r \phi \, d\phi \tag{5-4}$$

where $s = \sum_i n_i G_i P_i$ is the total effective radiated power per unit area.

If we take into account the spectral spread of the transmitted signals and

the selectivity characteristic of the receiver by assuming both are gaussian (that is, the transmitted signals are pulses of gaussian envelope, T seconds long, 1.086 decibels *down*, and that the receiver passband is b cycles per second wide, 1.086 decibels *up* from the minimum threshold P_r), the expected signal density becomes

$$N = \frac{\lambda^2 s b}{16 P_r B} \cdot \frac{bT}{\sqrt{\pi(4\pi^2 b^2 T^2 + 1)}} \int_0^{2\pi} G_r(\phi)\, d\phi \qquad (5\text{-}5)$$

An Extension to Transmission Beyond the Horizon. When the range of the receiver is likely to be limited by signal attenuation due to tropospheric scattering the expected number of signals is given by

$$N = \sqrt{5}\left(\frac{\pi}{2}\right)^{7/10}\left(1 + \frac{1}{4\pi^2 b^2 T^2}\right)^{1/10}\left(\frac{0.205 k m^8 \lambda^2}{2 P_r}\right)^{2/5}$$

$$\times \int_0^{2\pi} [G_r(\phi)]^{2/5}\frac{d\phi}{2\pi} \cdot \frac{b}{B}\sum_i n_i\,(G_i P_i)^{2/5}$$

$$+ \pi^{11/10}\left(1 + \frac{1}{4\pi^2 b^2 T^2}\right)^{3/10}\left(\frac{0.41 k m^8 \lambda^2}{P_r}\right)^{1/5}$$

$$\times \int_0^{2\pi} [G_r(\phi)]^{1/5}\frac{d\phi}{2\pi} \cdot \frac{b}{B}\sum_i n_i\,(G_i P_i)^{1/5} \qquad (5\text{-}6)$$

In both cases the actual number of signals will have a Poisson distribution with parameter N.

If it is not known in advance whether the furthest transmitter that can be picked up is beyond the horizon, both Eq (5-5) and (5-6) may be computed. The smaller of the two results should then be used since it corresponds to the range-limiting phenomenon that sets in first.

5.2.3 A Stochastic Process Model of the Signal Environment at the Output* of an Intercept Receiver

If the output of an intercept receiver as it sweeps in azimuth and searches in frequency is presented on a panoramic display, signals will appear to

*See Reference 7.

come and go. For a tactical situation where a wide variety of radiating equipment is in use it is reasonable to assume that the arrival and departures of signals may be described by a random process. Proceeding from this assumption a technique for relating parameters of interest in the design of pulsed signal-analyzing equipment to the expected number of signals arriving in some instant at the output of the intercept receiver is developed.

The point of view taken in this section is different from that in the previous sections. Signal environment now is considered to be a function of the receiver parameters as well as the geographic distribution of the transmitters. Those signals that could be distinguishable above a given noise level at the output of the intercept receiver are the only ones included. The signal density at some instant is the number of signals that arrive at the output of the receiver during some given short period of time—a period that is long with respect to the smallest pulse repetition period but short with respect to the sweep time in azimuth and frequency.

Definition of the Categories. In order to describe the signal environment we define four categories of signals in terms of their arrival characteristics at the output of the receiver. If we consider a signal whose power density at the receiver input is such that it is detected only when either the transmitter antenna points at the receiver or the receiver antenna points at the transmitter as two independent signals the following categories are mutually exclusive:

Category 1:   Signals that are received continuously at some fixed frequency regardless of the orientation of the receiver or the transmitter antenna.

Category 2:   Signals that are received at a fixed frequency when and only when the line joining the receiver and transmitter is contained in the beamwidth of the receiving antenna.

Category 3:   Signals that are received at a fixed frequency when and only when the line joining the receiver and transmitter is contained in the beamwidth of the transmitting antenna.

Category 4:   Signals that are received when and only when both the antenna beamwidth of the receiver and the antenna beamwidth of the transmitter contain the line joining the transmitter and receiver.

The signal density at the output of the receiver is described by writing the probability density function for each of the categories in terms of these system parameters: receiver bandwidth, antenna beamwidth, azimuth sector scanned by the receiver antenna, frequency region searched by the receiver,

transmitter antenna beamwidth, and the number of transmitters in the region. The number of signals available in each category is assumed to be given by a map of the geographic distribution of transmitter locations.

The Signal Density Model. On the assumption that the signals arriving at the output of the receiver, regardless of category, are uniformly and independently distributed over the frequency spectrum and azimuth sector of interest, probability density functions were obtained for each category. If we let X_i = the random variable ranging over the number of signals from the ith category received in a given instant, $(i = 1, 2, 3, 4)$, the number of signals at the output of the intercept receiver is described by the Poisson density function

$$P\left\{X_1 + X_2 + X_3 + X_4 = n\right\} = \frac{\lambda^n e^{-\lambda}}{n!}$$

The expected number of signals received at any instant is

$$\lambda = \left[N_1 + N_2 \frac{a}{A} + N_3 \left(\frac{\bar{a}}{A}\right)_3 + N_4 \left(\frac{\bar{a}}{A}\right)_4 \right] \frac{\alpha}{D}$$

where N_i — number of signals in category i $(i = 1, 2, 3, 4)$ for a given geographic distribution of transmitters,

a — receiver antenna beamwidth,

A — azimuth sector swept by the receiver antenna,

$\left(\frac{a}{A}\right)_j$ — an average probability and an estimate of the probability that a particular transmitter antenna beam of category j $(j = 3, 4)$ contains the line joining the transmitter and receiver,

α — apparent receiver bandwidth and

D — frequency spectrum searched by the receiver.

It is recognized that many of the approximations used in developing this model assume ideal system parameters and field situations. No attempt has been made to include signals from spurious responses in the receiver or to modify the N_i by using probability density loss functions that describe the on and off times for the transmitter. It is hoped that this model sheds some light on how the system parameters used here affect the signal density. Perhaps data from future field tests can suggest modifications to this model that will make it more realistic.

To apply these techniques to the Rand Corporation maps published in

Rand-280, assume that we are using a receiver with sensitivity 85 decibels below a milliwatt, a gain G_r in the main beam of 30 decibels in a 5-degree beamwidth, an average gain on the back and side lobes of 7 decibels, and an effective bandwidth of 10 megacycles per second. These enable one to estimate the N_t. Also assume we are searching a frequency band of $D = 2000$ megacycles per second, essentially the S-band, and an azimuth sector of $A = 90$ degree in which the typical transmitters have an average gain on the back and side lobes of $G_t = -10$ decibels and a gain in the main beam that is greater than the gain in the main beams of the receiver. These assumptions and the assumption that all transmitters are turned on, lead to the conclusion that 95 percent of the time one would expect no more than two signals at the output of the receiver at any given time.

Preliminary studies aimed at comparing the results of the Monte Carlo model and the stochastic process model suggest that perhaps the Monte Carlo method counts the signals from one transmitter too often. In the other approach the basic assumption of a uniformly random distribution of transmitters is actually violated and thus its result gives too low an estimate of the signal density in some regions.

5.8 An Application of Pulse Counting to Obtain Tactical Electronic Intelligence Information

This section considers a brighter side to the signal environment problem. The complex signal environment in a field army area can be used to supply useful intelligence information. The device used to gather this information is a simple pulse counter rather than a complex radar "fingerprinting" device.

5.8.1 System Philosophy

The system decided upon as meeting the requirements for a tactical ELINT sensing device may be summarized as follows: Three low-sensitivity receivers are installed in a drone along with pulse counting and recording devices. The drone is then flown on a preassigned path which, based upon receiver sensitivity, will cover the entire front line and part of the communication zone of the enemy army. The system "counts" the number of pulses received and records the results as a function of time. This information is recovered after the drone lands, or by broadcast to a receiving site in friendly territory where the density is read out and plotted on maps of the tactical area. The contours resulting may then be used in intelligence inference.

Basic to the philosophy is the assumption that the enemy will defend his most important formations most heavily.

5.8.2 System Characteristics

Any number of drones exist capable of carrying out the payload necessary

to accomplish the mission. The drone used for the example discussed here is the RB-77, now under development. The drone should fly as low as possible, consistent with its safety. A low-flying aircraft is difficult to track and prevents the use of atomic warheads in surface-to-air missiles. From a data gathering standpoint, it prevents several antiaircraft systems from simultaneously tracking the drone and thus disturbing the pulse counting.

Within the drone the receiving and recording system should cover the frequency region of the major enemy radiating equipments. A possible con-
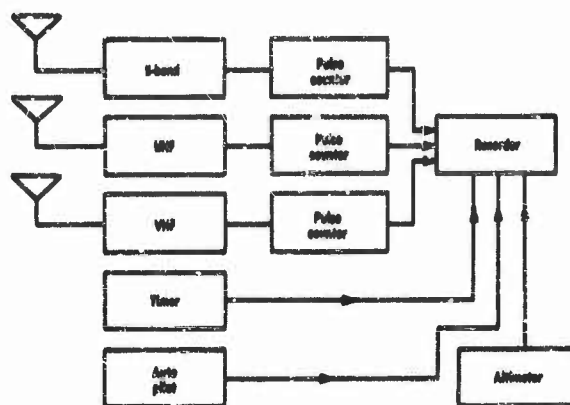
FIGURE 5-3   Block Diagram of Electronic Equipment

figuration of receiving and recording equipment is shown in Figure 5-3. The autopilot commands should be recorded in order to know the track of the drone. The altimeter reading should be recorded to help resolve ambiguities and to determine effects of terrain on the data whenever this is necessary.

Continuous rebroadcast of data may compromise the system and simplify passive tracking by the enemy. To prevent large scale spoofing the purpose of the drone flights should be carefully guarded. To insure a high probability of survival of the drone, enemy tracking should not be facilitated. Therefore, the system should record the data gathered and either rebroadcast it in occasional short spurts or store it for analysis after the drone lands.

5.3.3   Results of a War Game Test of the Philosophy

In order to test the feasibility of the scheme suggested, a war gaming technique or sampling experiment is needed. Since a Haller, Raymond and Brown (HRB-Singer) report (Reference 6) was readily available, it was decided to use the development given therein to test the model. Certain assumptions were made which are given below:

Antenna Assumption: The actual form of the Whiff, Firecan and Cross-fork antennas probably appears somewhat as in Figure 5-4a. The assumed antenna pattern for use in the game is shown in Figure 5-4b; that is, we
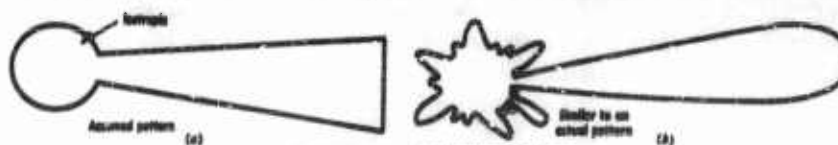
Figure 5-4 Antenna Pattern

assume an isotropic pattern except for the main lobe of each antenna. This is a common assumption in preliminary studies of this type.

Sensitivity Assumption: It was very easy to obtain a sensitivity in all three receivers to give a four-mile range against the various enemy equipment, when the receiver is in the fine structure of back and side lobes (assumed isotropic) of the equipments. Thus, it was assumed that the sensitivity of each receiver was set at a level such that a signal was recognized, whenever the drone was within four miles of the transmitter source regardless of the transmitting antenna's orientation. Of course, a signal would be recognized and counted whenever the drone was within the major lobe of the source equipment antenna at much greater distances.

The actual data gathering was carried out in a manner described in this section.

First, the HRB-Singer report was used to provide a tactical model of the field army engaged in battle over several days. Searches with the drone occurred at 2200 on 16 June, 1700 on 17 June, and 1200 on 17 June. These times were selected arbitrarily since they were the first three maps in the HRB-Singer report. It would probably be desirable to search more frequently in an actual battle to keep closer contact with enemy movements. Data reflected the existence of Crossfork, Whiff, and Firecan.

Figure 5-5 which shows the S-band density found at 2200 on 16 June, will be used in the discussion. The other five contour overlays are not shown.

A flight path was selected based upon drone characteristics (in this case the RB-77) and receiver sensitivity. The flight path for the 2200 flight is shown as a dashed line. The marks are miles flown.

Next a circle of four-mile radius was moved with its center constrained to lie on the flight path line. The number of radars in the circle were counted. The count represented the density of radar signals received at the drone which was simulated to be at the center of the circle. The equipment would actually record the number of pulses per second or pulse rate. Using this data, a contour map was drawn as shown. The information was then in suitable form for use by an intelligence officer. Although not shown, the same pro-
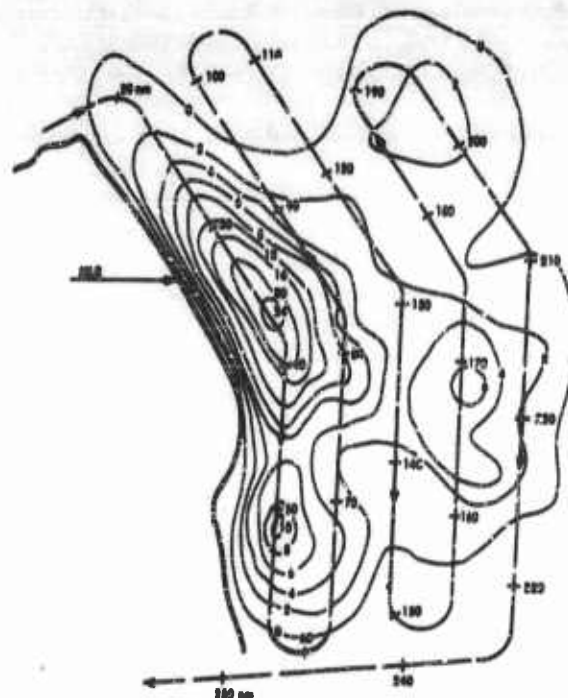
FIGURE 5-5   Signal Density Contour Map

cedure was done at each of the times mentioned above and in the UHF and S bands. This was done to see how the concentrations changed.

In making these contour maps, actual pulse count information is know.. along the flight path only. A reasonable uniformity of the underlying distribution of pulse counts is then assumed so that the contours may be added by interpolating between points from two parallel flight paths. If there is some reason to doubt the validity of such an assumption on any particular mission, an additional flight can be used to give more detailed information.

5.3.4 Random Elements

In the field application of this technique, many random errors will be introduced. The most troublesome may be the flight path errors. Since the drone is not tracked and no attempt is made to correct during flight for the effects of air turbulence there may be sizeable deviation from the assigned path. If meteorological data indicate steady winds, the magnitude of these errors may be reduced by accounting for the predicted wind in the autopilot program for each leg of the flight path. A rough fix on the magnitude of the

residual error can be obtained from a comparison at flight termination of the programmed landing area and the actual launching area. In general when the air is turbulent, flights should be short to keep errors within tolerable limits.

Other types of errors such as counting errors due to tracking of the drone by enemy radar from time to time, reflections from prominent terrain features or just propagation anomalies should not have great effect on the inferences drawn from the contour map. The major conclusions are based on relative magnitudes of the pulse counts rather than the absolute count. If doubt exists about certain regions on the map, additional flights may be used.

For each of the above errors, attempts could be made to design remedies into the equipment. We feel, however, that such an approach is not compatible with the constraints mentioned in the beginning. Intelligent adjustments of operating procedures can readily handle most situations which may arise. In using this sampling technique as the experience of the operator increases, the information gathered becomes more complete and more reliable. Training and experience with these simple techniques will contribute much more than complexity added in an attempt to anticipate every problem.

5.3.5 Intelligence Information

Based on information gained from the contour maps it was possible to make many inferences regarding the enemy's activities. As with any intelligence effort, skill develops with extensive use. Hence, the equipment suggested would yield far more information as the users become more skilled in interpreting the concentrations and changes. However, even without experience, it has been possible to infer information of the following type:

(a) Where the enemy would attempt the crossing was clear from the first map shown in Figure 5-5. If information on how rapidly this build-up was occurring was available in form of similar contour for 1200, 1600, and 1800, the same day, it would have been possible, based on experience with previous attacks to infer quite closely the time the attack would occur. Even the 2200 map indicates the time is close at hand and doctrine implies the attack is probably scheduled for early the next morning.

(b) Major supply areas, depots and railheads appear as more permanent defenses. Even secondary railheads appear.

(c) Concentrations of troops of secondary size show up. This would indicate a feint or lesser magnitude attack from that direction.

(d) In each of the above cases, an indication of the size of the atomic weapon needed to disrupt the enemy plans was indicated. Also, the magnitude of the air defense indicates the use of aircraft artillery or ballistic missile as a delivery means.

(e) Any shift in air defense can be detected when the enemy is moving supply depots forward to new locations. This in turn indicates the time for interdiction fire.

Analysis by experienced personnel and more frequent flights would undoubtedly increase the number and types of conclusions which could be drawn from these surveys.

5.4 Conclusion

Most of the effort in signal environment study has been directed toward describing the environment in a tactical situation. From the resulting model one attempts to deduce the type of interference problem that may arise or to simply describe the signal density at a receiver output located in the environment. The models obtained to date are by enlarge too simple or too rigid to lend themselves to the variety of problems that involve such information. The variety of results obtained at many different laboratories throughout this country constitutes a good start on the general problem. There is a continuing need for more effort to be expended.

An example of an area where effort is just beginning arises in evaluation of several operational ELINT systems. The signal environments at their operational sites need to be adequately described. A flexible, rapid and reasonably detailed method for analyzing these environments is not available from the work done on tactical situations. Digital computer simulation methods for application to different phases of the over-all problem are being formulated at several laboratories. Many simplifying decisions must be made to bring this complex problem into a tractable form for the IBM704. Does oversimplification offer the most meaningful results? How can the program be set up to allow additional refinements as the information becomes available? Perhaps a group of semi-independent subroutines which can be modified with a minimum of effort and whose results can be readily combined into an over-all meaningful description of the environment will be developed.

Many powerful mathematical and statistical tools have been developed in very recent years. As these reach the applied efforts new and different approaches to the problem will appear. Better techniques for more precise descriptions of the environment and for design and operation of electronic systems to make them more independent of the signal environment are badly needed.

REFERENCES

1. Hiebert, A. L., "Signal Density Surface, Radar Development," Rand Report R-280, (1955) (SECRET)

2. Signal Corps Engineering Laboratories, "Final Report, Project MONMOUTH I," Vol. II, January 1956 (SECRET)

3. Myers, H. A. and Van Every, A., "Prediction and Measurement of Radar Signal Densities," Rand RM-1700 (SECRET)

4. Melpar, Inc., "Study of the Electromagnetic Radiation Complex in the Battle Area," (SECRET)

5. Harley, J., "The Tactical Signal Density," Sylvania Electronics Defense Laboratory, EDL-82, (SECRET)

6. Blachman, N., "On the Expected Number of Signals Picked up Simultaneous by a UHF Receiver," Sylvania Electronics Defense Laboratory, EDL-M53, 1 September 1955, (CONFIDENTIAL)

7. Fossum, R. and Halteman, A. E., "The Estimation of the Signal Density in a Tactical Situation," Sylvania Electronics Defense Laboratory, (Unpublished report) (SEC

8. "DETKIV-ASA", Haller, Raymond and Brown (HRB-Singer), 60.7-R-4, (1957) (SECRET)

6

Intercept Probability and Receiver Parameters

A. B. MACNEE,   D. B. HARRIS

6.1 Introduction

Electronic warfare may be analyzed as a game. For every electronic device that a potential opponent can employ there always exists a countermeasure which will reduce its effectiveness. The feasibility of applying electronic countermeasures, however, depends to a marked degree upon the state of our knowledge concerning the enemy's electronic systems. Two classes of information are important. The first of these is what might be termed strategic information; it includes such things as technical characteristics of the system to be countered, the mode of operation of the system, and the nature of supplementary systems which can be employed by the enemy. This type of information is needed to make strategic decisions such as whether one should attempt to counter the system and what characteristics are required for the countermeasures device. The second class of information is tactical: Is the enemy using a certain electronic system? What frequency is he on? Has the enemy shifted his frequency as the result of our jamming signal? These are examples of questions which must be answered in the field if electronic countermeasures are to be successfully employed.

One of the most important sources of both strategic and tactical intelligence concerning a potential enemy's operations is the interception and analysis of the signals radiated by his electronic systems. Clearly, before one can perform any sort of analysis of signals, the signals must be received. The

important initial problem in attempting to gather electronic intelligence is, if an enemy system is operating and is generating electromagnetic radiation, what is the probability of this radiation being detected and recognised? This may be defined as the probability of intercept for a given electromagnetic radiation. Clearly this probability is a function of the parameters of both the signal and the receiver employed. The objective of this chapter is to investigate the problem of calculating or estimating this probability of intercept. Although several specific signals are considered for illustrative purposes, the emphasis is on general approaches to the problem and on general trends which apply to a broad class of signals.

6.1.1 Signal Detection

In considering signals which an intercept system may encounter it is illuminating to classify them according to parameters such as: center frequency of radiated signal, spectral width of signal, signal duration, signal waveform, signal strength, etc. If the form of a signal is known exactly, a body of signal detection theory is available to predict the best that can be done in detecting its presence when masked by additive gaussian noise (References 1, 2, 3, 4). In such a case the probability of intercepting such a signal can be reduced to the probability of detecting it in the presence of noise. This probability is a function of the probability of a false alarm which one is willing to tolerate and the ratio of the signal energy to the noise power per unit bandwidth. This problem is treated in Chapter 7. The optimum fixed-observation-time receiver performance is summarized by the receiver operating characteristics which are plotted in Figure 6-1. $P_{SN}(A)$ is the probability that if a signal is present in the noise, it will be detected. $P_N(A)$ is the probability that if noise alone is present, it will be reported as a signal; this is a false alarm. The receiver operating characteristic (ROC curve) plots $P_{SN}(A)$ versus $P_N(A)$ with the detectability index d as a parameter. The detectability index is the difference between the means of the probability density functions $f_{SN}(x)$ and $f_N(x)$ divided by

FIGURE 6-1 Receiver operating characteristic for a signal known exactly. $\ln[l(x)]$ is a normal deviate, $\sigma_{SN}^2 = \sigma_N^2$, $(M_{SN} - M_N)^2 = d\sigma_N^2$. σ_{SN} and M_{SN} are the standard deviation and the mean, respectively, of $f_{SN}(x)$; σ_N and M_N are the standard deviation and the mean, respectively, of $f_N(x)$.

the variance of $f_N(x)$. The curves of Figure 6-1 are for a signal having an exactly known waveform (only its presence is uncertain) and for this case

$$d = 2E/N_0 \qquad (6\text{-}1)$$

E is the signal energy received during the observation interval and N_0 is the noise power per unit bandwidth. For a fixed d the ROC curves are seen to be straight lines with unity slopes when \. ed on double probability paper (Figure 6-1). In this ideal receiver a signal is reported whenever the likelihood ratio.

$$l(x) = f_{SN}(x)/f_N(x) \qquad (6\text{-}2)$$

exceeds some threshold level β, where $f_{SN}(x)$ and $f_N(x)$ are the probability density functions for a given output x when signal plus noise or noise alone are present at the input.

$$P_{SN}(A) = \int_{-\infty}^{A} f_{SN}(x)\ dx \qquad (6\text{-}3)$$

and

$$P_N(A) = \int_{-\infty}^{A} f_N(x)\ dx \ . \qquad (6\text{-}4)$$

The curves of Figure 6-1 apply to any case in which

$$\ln [l(x)] = \ln [f_{SN}(x)] - \ln [f_N(x)] \qquad (6\text{-}5)$$

is normally distributed with the same variance both with noise alone and with signal plus noise. The curve labeled $d = 0$ corresponds to zero signal energy. In this case, no matter what threshold level β is chosen, the probability of a "detection" is the same as the probability of a false alarm. In other words, the presence of a signal of zero energy has no influence on the receiver performance; in this case, one may just as well throw the receiver away and flip a coin.

These curves for the case of a signal which is exactly specified are of particular importance in that they set an upper bound for any practical situation where some additional uncertainty is bound to exist. The ROC curve is a convenient form in which to present data on a receiver; it can be used

to generate other forms of the receiver's performance characteristics. If one wishes to plot the probability of a correct detection for a fixed false alarm rate as a function of signal intensity one would take a vertical cut on Figure 6-1 at the specified false alarm rate. Thus, if the signal is a 0.5 microsecond pulse of S watts peak power masked by gaussian noise having a uniform power density of N watts per cycle per second over a bandwidth of 1 megacycle per second, $d = S/N$;* and for a false alarm probability of .001 the probability of a correct detection varies as sketched in Figure 6-2 with $2E/N_0$ in decibels. For this rather low false alarm rate one sees that a signal-to-noise ratio of 13 decibels is required to give a detection probability of 90 percent on a single trial. If one had instead ten pulses of the same peak power available this would raise the signal energy by a factor of 10 so that $2E/N_0 = 10(S/N)$. The signal to noise power ratio necessary to give a 90 percent probability detection with the same false alarm probability would then be reduced to 3 decibels.

6.1.2 Effects of Signal Uncertainty

Figures 6-1 and 6-2 are both calculated on the assumption that everything
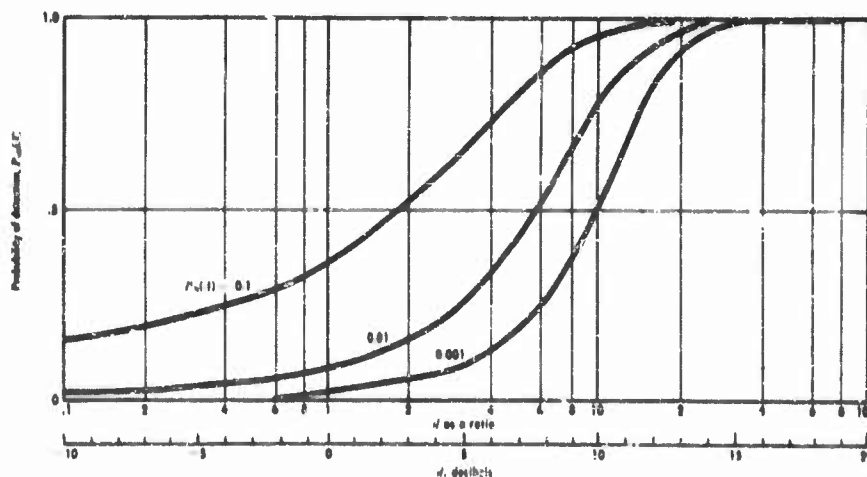
FIGURE 6-2   Probability of a correct detection of a signal known exactly, versus d.

about the signal is known including the value of d for the signal. The only uncertainty is as to whether the signal occurs in the specified time interval or not. Whenever there is any additional uncertainty concerning the signal,

*$2E = 2S \times \tau$ joules and $N_0 = N/B$ watts per cycle per second, therefore $2E/N_0 = d = (S/N) \times 2B\tau = (S/N) \times 2 \times 10^6 \times (1/2) \times 10^{-6} = S/N$.

the ideal receiver performance is poorer than that predicted in these two figures. Considerable effort has been expended in studying cases where one's knowledge of the signal expected is uncertain in some fashion. Peterson and Birdsall consider the cases listed below (Reference 5).

1. A signal known except for phase,
2. A signal which is a sample of white gaussian noise,
3. The video output pulse of a broadband receiver of known starting time,
4. A signal which is one of M orthogonal signals, all known exactly,
5. A signal which is one of M orthogonal signals known except for phase.

Case 3 is of considerable practical interest since it can be applied directly to the case of a broadband intercept receiver looking for a pulse signal of unknown center frequency. If the bandwidth of the wideband receiver is W cycles per second, and if the signal duration is M/W seconds, then for weak signals

$$(1/M) \cdot (2E/N_0) < < 1 \tag{6-6}$$

the receiver operating characteristics of Figure 6-1 apply (see Section 4.5 of Reference 5) with

$$d = (1/4M) \ (2E/N_0)^2 \tag{6-7}$$

This represents a very considerable degradation of performance from the case of a signal known exactly. If, for example, one is looking for a pulse of 0.5 microsecond duration but unknown center frequency in the frequency range from 2800 to 3200 megacycles per second, one has $M = 300$. To obtain a detectability index of 9 would require

$$(2E/N_0) = \sqrt{800 \times 9} = 84.7,$$

which is 9.73 decibels below the performance possible if the signal were known exactly.

The broadband amplifier and detector followed by an optimum video amplifier is not the best way to detect a signal of unknown center frequency. Case 4 listed above can be applied to estimate the optimum performance if the signal is known exactly except for its center frequency. The ideal receiver for this case is one which splits the frequency band into M channels. The output of each channel is cross-correlated with one of the M exactly

specified waveforms. The correlator outputs are then combined with an exponential weighting which depends upon the known noise level. The exact evaluation of the performance of this receiver is very difficult to obtain, but Peterson and Birdsall show that assuming that the logarithm of the distribution of the likelihood ratio is normal, the performance is approximately that shown in Figure 6-1 (see Section 4.8 of Reference 5) with

$$d = \ln\left[1 - \frac{1}{M} + \frac{1}{M}\exp(2E/N_0)\right]. \tag{6-8}$$

The improvement possible with this much more complicated receiver is considerable. Using the same numbers considered for the broadband video example, one finds for a detectability index of 9

$$(2E/N_0) = \ln[1 + M(e^d - 1)]$$
$$= \ln[1 + 200(e^9 - 1)] = 14.3. \tag{6-9}$$

This is only 2.01 decibels below the performance possible when the signal is known exactly and is 7.73 decibels better than the wideband amplifier followed by an optimum video amplifier.*

Equation (6-8) gives the detectability index that can be achieved by an ideal receiver looking for one of M signals all specified exactly. Case 5 above treats the case of the M signals known exactly except for the carrier phase. For this case the optimum receiver can achieve a detectability index of approximately

$$d = \ln\left[1 - \frac{1}{M} + \frac{1}{M}I_0\left(\frac{2E}{N_0}\right)\right] \tag{6-10}$$

where I_0 is the Bessel function of zero order and purely imaginary argument. For $2E/N_0$ greater than 4

$$I_0(2E/N_0) \approx \sqrt{(N_0/4\pi E)}\exp(2E/N_0) \tag{6-11}$$

to an accuracy of better than 4 percent. Solving for the ratio $2E/N_0$ necessary to achieve a given detectability index gives

$$\ln\left[I_0\left(\frac{2E}{N_0}\right)\right] = \ln[1 + M(e^d - 1)] \approx \frac{2E}{N_0} - \frac{1}{2}\ln\left(\frac{4\pi E}{N_0}\right). \tag{6-12}$$

*Note this comparison does not include any possible difference in the amplifier noise factors for the two systems.

Considering again the conditions of $M = 200$ and $d = 9$, then from Eq. (6-12) the necessary ratio of signal energy to noise power per unit bandwidth is 16.6. This is only 0.66 decibels more than is required if the phase were known and 2.67 decibels more than what is needed if the signal is known exactly.

In this section an attempt has been made to give the reader some feeling as to the optimum possible performance in detecting an exactly specified signal, when that signal is masked by white gaussian noise (Eq. 6-1 and Figure 6-1). Equations (6-9) and (6-12) give a measure of the minimum degradation that is introduced into the detection performance by the addition of a limited amount of uncertainty with regard to the signal, the signal being one of M orthogonal signals all known exactly or all known exactly except for the carrier phase. Finally Eq. (6-7) gives a measure of the much greater degradation in performance that will occur if this uncertainty is met by the simplest means, just broadbanding the receiver. This comparison is summarized in Figures 6-3 and 6-4 which plot the increase in signal energy
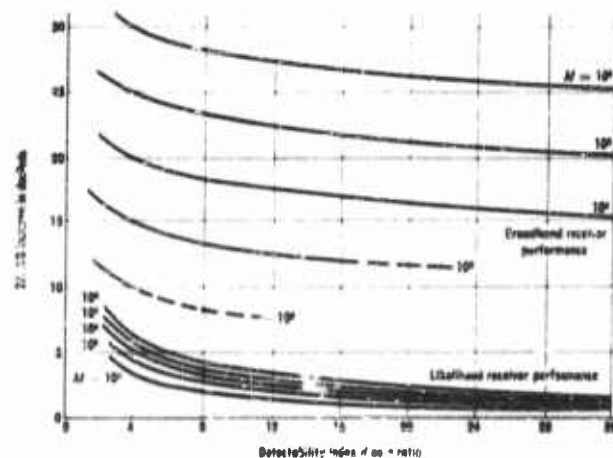
FIGURE 6-3 Comparison of increase in signal energy necessary to maintain a fixed detectability index as signal uncertainty is introduced. M = number of orthogonal signals.

in decibels relative to the zero uncertainty case (signal known exactly) as a function of the detectability index d; the number of orthogonal signals M, is the parameter of these curves. Figure 6-3 compares the ideal likelihood receiver for the case of all M signals known exactly with the simple broadband receiver employing a square-law detector and an optimum video filter. Grossly, the degradation in performance is very large when the degree of
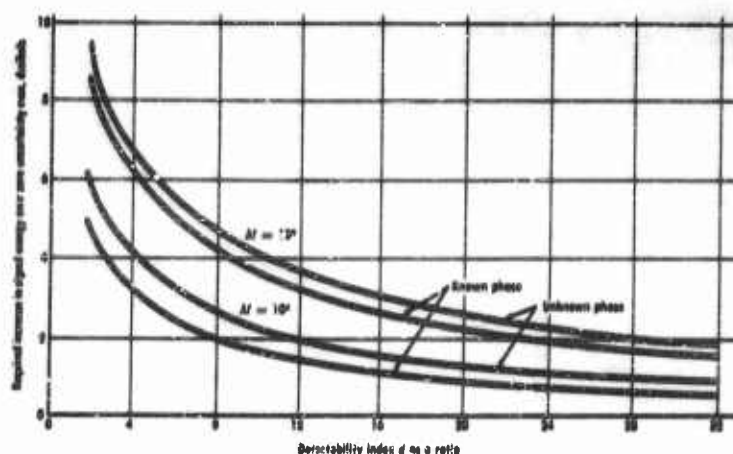
FIGURE 6-4 Effect of signal uncertainty on signal detectability, M orthogonal signals.

uncertainty is high and the simple, broadband receiver is used. The use of an optimum likelihood ratio receiver can reduce this degradation very significantly particularly if one is interested in the larger values of d. Referring to Figure 6-1, one sees that this corresponds to being interested in signals such that the false alarm probability can be kept low and the detection probability high. Figure 6-4 demonstrates that if one can use a likelihood receiver the degradation in performance to be expected if the signal phase is not known is not very great ($0.3 \sim 1.25$ db for $10^2 \leq M \leq 10^6$).

These curves give a measure of the best one can hope to do in detecting the presence of a signal under practicable conditions. Normally one does not deal with orthogonal signals nor with signals specified exactly or exactly except for phase. The additional uncertainties present in a real situation can be expected to result in performance below that predicted in the figures. This loss in performance can be anticipated to be more significant for the likelihood receivers than in the case of the broadband receiver, since the latter make more use of one's knowledge of the possible signals which may occur.

Despite these weaknesses, the curves of Figures 6-3 and 6-4 have utility in estimating or judging receiver performance. As an example one might consider the problem of detecting the presence of a pulse signal from a radar of known characteristics. The radar signal is known to have a duration of one microsecond, a repetition rate of 1000 pulses per second, and to be in a frequency band 100 megacycles per second wide. Let us suppose we are interested in detecting the radar signal when it is looking at our receiving

antenna, but that the direction to the radar from the receiver is unknown. To arrive at an estimate for the parameter M in this case one can assume that every radar pulse is to be detected. This means that one thousand times a second (at the end of every millisecond) the receiver must make the decision; a pulse was or was not received during the past millisecond. Since the one microsecond pulse can be anywhere in each one thousand microsecond interval, an estimate of this uncertainty in arrival time for the pulse is a factor of 1000. The spectrum occupied by a single pulse is about one megacycle per second wide, and it can lie anywhere in the band 100 megacycles per second wide. This frequency uncertainty combined with the time of arrival uncertainty gives 10^5 possible signals. Figure 6-1 shows that a detectability index $d = 16$ allows one to achieve an 81 percent detection probability with a false alarm probability of 0.1 percent. If a broadband receiver is used, then, from Figure 6-3, $2E/N_0$ must be 22 decibels above 16, or 2560. Since

$$E = S \times t_d \tag{6-13}$$

where $S =$ peak pulse power
    $t_d =$ pulse length

and

$$N_0 = FhT \tag{6-14}$$

where $F =$ receiver noise factor
    $T =$ temperature seen by receiving antenna
    $h =$ Maxwell-Boltzmann constant $= 1.37 \times 10^{-23}$ joule

one can solve for the peak signal power necessary to give the detectability index:

$$S = (FhT/2t_d)(2E/N_0). \tag{6-15}$$

Assuming a noise factor of 10 and a temperature of 300 degrees Kelvin*,

$$S = \frac{10 \times 1.37 \times 10^{-23} \times 3 \times 10^2}{3 \times 10^{-6}} \times 2.56 \times 10^3$$

$$= 52.6 \times 10^{-12} \text{ watt.}$$

It should be noted that the assumed false alarm probability of 0.1 percent

*A reasonable figure if antenna looks along the surface of the earth.

will lead to the rather high false alarm rate of 1 per second since 1000 decisions are made per second. If a false alarm rate of 1 per hour is desired the false alarm probability on each trial must be reduced to 2.78×10^{-7}. To achieve a detection probability of 0.90 with this false alarm probability requires an increase in the detectability index from 16 to 41.3. If one estimates from Figure 6-3 that at this value of d the broadband receiver loss is 20 decibels, the required peak signal power from Eq (6-15) becomes 84.8×10^{-12} watt.

A block diagram of the broadband receiver evaluated above is given in Figure 6-5(a). The output of the square-law detector is fed to a gated integrator. At the end of each millisecond the output of this integrator is clamped to zero volts. The output of the integrator is compared with an adjustable threshold voltage; and whenever the integrator output exceeds this voltage, a detection is indicated (by lighting a light, ringing a bell, etc.).

From Figures 6-3 and 6-4 one sees that a reduction of about 20 decibels in the signal energy required at the receiver input to give the same detectability can be achieved through the use of a likelihood receiver. This is a large improvement; it would permit an increase in the free space intercept range of a given signal of ten times! Figure 6-5 (b) gives a block diagram of this receiver. The cost of this receiver designed to give optimum performance in the face of 10^5 possible signal waveforms is seen to be prohibitive. After separating input passband into one hundred separate passbands, corresponding to the possible pulse center frequencies, it is still necessary to cross correlate the output of each narrow filter with the one thousand possible signal waveforms at each center frequency. These correspond to the one thousand possible true positions of the signal pulse. The outputs of the 10^5 correlators are then given an exponential weighting and combined in a single adder. The output of this adder is the likelihood ratio, and it would be compared with an adjustable threshold voltage in a voltage comparator circuit to give the detection indication. This astronomical in-

FIGURE 6-5 Block diagram of receivers for the detection of a pulse signal of unknown center frequency and starting time. (a) Broadband receiver. (b) Likelihood ratio receiver.

crease in equipment complexity necessary to optimally detect a signal when the degree of uncertainty concerning the signal is large has forced a variety of engineering compromises on those who are concerned with the design and development of practical intercept equipments.

It should be noted, however, that the complex receiver of Figure 6-5(b) in addition to doing an optimum job in the detection of the presence of a signal also gives one a relatively reliable estimate of characteristics of the signal received. This can be obtained by comparing the outputs of the cross-correlators whenever a detection occurs. The correlator having the largest output should indicate which signal caused the detection. Birdsall and Peterson have investigated the probability that this largest output will be the correct choice among M alternative choices if a signal has occurred (References 6 and 7). This probability depends upon the detectibility index d. Let this probability be denoted $P_M(d)$. Birdsall shows that

$$P_M(d) = \Phi(a_M d^M - b_M), \qquad (6\text{-}16)$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\,(t^2/2)\,dt, \qquad (6\text{-}17)$$

$$\Phi(-b_M) = 1/M, \qquad (6\text{-}18)$$

and

$$a_M = \alpha_M/1.28255 \qquad (6\text{-}19)$$

For $M > 1000$, a_M can be obtained from the equation

$$\Phi(a_M) \approx 1 - (1/M) \qquad (6\text{-}20)$$

Values of a_M for several values of M are tabulated in Table I. In general, for large values of M and for large detectability indices, $P_M(d)$ will be very close to unity.

TABLE I

M	4	16	256	10^5	10^6	10^9	10^{12}
a_M	.827	.884	.916	.964	.968	.979	.984

The receiver of Figure 3-5(b) can be simplified considerably by the use of a gated matched filter instead of the parallel cross correlators.* The output of a matched filter at the end of each possible pulse position will be the cross correlation between the input signal and pulse at that time. This output can be sampled and then applied through an exponential function generator to a summing integrator together with the gated output of the other ninety-nine matched filters. Such a likelihood receiver is indicated in block diagram form in Figure 6-6. This receiver has the same detection capability

as the receiver in Figure 6-5(b), but the signal recognition capability has been lost. It could be recovered by the introduction of suitable memory and voltage comparator circuits at the outputs of the matched filters.

If the pulse position is not known to be at one of 1000 nonoverlapping positions, then the pulse generator and gate circuits of Figure 6-6 should be removed. The matched filters would be passed through

FIGURE 6-6 A simplified likelihood ratio receiver for 10^2 orthogonal signals.

outputs of the one hundred matched filters would be passed through exponential function units and then just summed in a single integrator. The output of this integrator would then be sampled at the end of each repetition period. This simplified receiver is believed to be optimal for the detection of a pulse of unknown time position within the repetition period, but the performance of this receiver has not been calculated. Its performance can be expected to be poorer than that of the receiver in Figures 6-5(b) and 6-6, since more uncertainty has been introduced.

6.2 Approximate Approaches to the Intercept Problem

The previous section has introduced some of the problems of detection and recognition of signals when large degrees of uncertainty are involved. The complexity of this problem has forced the use of approximate methods for studying the problem of intercepting signals under practical conditions. In this section some of these approximate methods of analysis and the results obtained with them will be reviewed.

6.2.1 The Coincidence Concept

In considering the optimum detection of any one of a large ensemble of

*A matched filter is one whose impulse response is the expected signal waveform reversed in time. Thus, if the signal waveform is $s(t)$ and has a duration θ, the matched filter has an impulse response $s(\theta - t)$.

possible signals we have seen that the general approach is to build an optimum receiver for each possible signal. The outputs of these individual receivers or channels are then combined in a way which maximizes the difference between the output when a signal is present in at least one channel and the output when no signal is present. In such a receiver the way in which the probability of detection varies with signal strength is obtained by taking a vertical section through the receiver operating characteristics. Such a curve is plotted in Figure 6-7 with detectability index d as the abscissa. In general,
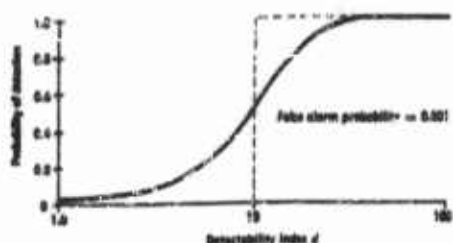
Figure 6-7   Probability of detection for a likelihood ratio receiver versus signal detectability index d.

d will be some function of the input signal-to-noise ratio as discussed in Sections 6.1.1 and 6.1.2. This figure is obtained by taking a vertical section through Figure 6-1 at $P_N(A)$ $= .001$. If a smaller false alarm rate is of interest, this curve would be shifted to the left, and vice versa. Probability of intercept calculations can be greatly simplified by replacing this detection curve by the dotted one shown. The assumption made by the dotted curve is that signal-to-noise ratios larger than some critical value (leading to $d = 9.5$ for the case shown in Figure 6-7) are certain to be detected, and all signals below this critical value are missed entirely. This approximation to the true situation has been called the "threshold concept," the critical value of d being the detection threshold. In Figure 6-7 this threshold value of d is chosen as the one leading to a 50 percent detection probability. This choice is arbitrary, and one might equally well choose a detection threshold corresponding to a 90 percent or greater detection probability.

Under the threshold assumption, any time the signal input to the receivers of Figure 6-5(b) or 6-6 falls within the acceptance band of one of the receiver channel amplifiers and has sufficient intensity, a signal interception will be made. On the other hand, if the input signal is rejected by all amplifiers or if its energy is too small, no interception occurs. Thus, with the threshold assumption, probability of intercept becomes the probability of coincidence between the signal parameters (field strength, center frequency, time of arrival, etc.) and the receiver parameters (threshold intensity, acceptance frequency range, etc.). Under this assumption the probability of intercept for all three receivers suggested at the end of Section 6.1.2 is unity for the intended radar signals as long as the input signal strength is sufficient. The threshold pulse energy for the two likelihood ratio receivers, however, will

be 20 decibels less than for the "broadband receiver" of Figure 6-5(a). Thus, calculation of the probability of intercept for any one of these receivers is reduced to calculating what fraction of the time the incoming pulse signals exceed the particular receiver's threshold energy.

As an example, let us suppose that the radar is being rotated at a uniform rate through 360 degrees of azimuth, and let it further be assumed that when the intercept receiver is within the radar antenna beamwidth the pulse energy exceeds the threshold of the likelihood ratio receiver by 35 decibels. If the radar side lobes are only down 30 decibels from the main beam, the intercept probability with likelihood receivers would be unity. The broadband receiver's threshold being 20 decibels lower, this receiver would only intercept pulses while the intercept receiver antenna was within the radar antenna's main beam. The probability of intercept for this receiver would thus be

$$P(I) = \theta_B/360° \tag{6-21}$$

where

$$\theta_B = \text{radar antenna bandwidth.}$$

6.2.2 Probability of Coincidence Calculations

Coincidence calculations have formed the basis of the intercept probability estimates by a number of authors who are listed in References 8 through 15. Two of the most recent reports in this area are those of Kiel and Enslow (References 13 and 15). Enslow's report is a rather careful and complete summary of the nature of the problem and of the prior work in the field. Such an exhaustive review does not seem feasible here, but some typical problems will be discussed to illustrate the approach.

In practical intercept systems the complexity of multichannel receivers has led engineers to the development of the scanning or panoramic receiver. One can consider the use of receivers that scan over any one of a variety of signal parameters such as: frequency, direction of arrival, intensity, duration, etc. Initially one is usually most interested in frequency and direction of arrival.

Figure 6-8 is an example of a time-frequency diagram for a receiver scanning the frequency band from f_1 to f_2 with linear sawtooth sweep. This figure can be used to calculate the probability of coincidence in both time and frequency between a pulse signal and the panoramic receiver acceptance frequency band. Figure 6-9 is an example of an angle-versus-time diagram for the visualization of coincidence between the directions of rotating re-
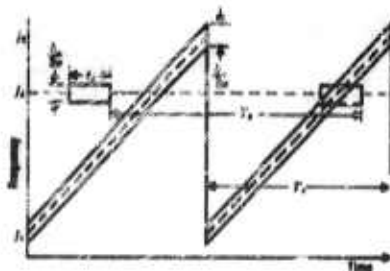
FIGURE 6-8 Frequency-t   diagram for a linearly sweeping panoramic receiver and a pulse signal.
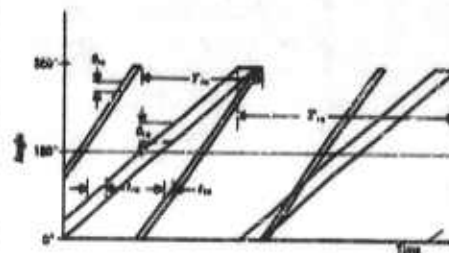
FIGURE 6-9 Diagram of antenna direction versus time for two narrowbeam antennas rotating at a constant speed.

ceiving and transmitting antennas and the angle of the path between the receiver and the transmitter. In this figure it is assumed that the reference directions for each antenna are chosen so that 180 degrees corresponds to both antennas "looking" directly at each other. Whenever the 180-degree line is within both beamwidths at the same time, an angular coincidence occurs. Such a coincidence is not illustrated in Figure 6-9, although it looks as though one might occur on the next rotation of the receiving antenna. The symbols used in Figures 6-8 and 6-9 are defined below. In choosing values for many of these quantities the threshold concept is used. For example, if the signal pulse duration is t_d seconds, we know that most of the signal power spectrum lies within a bandwidth approximately $1/t_d$ cycles per second wide. On the other hand, if the signal strength is sufficiently great, a receiver tuned to a frequency band $3/t_d$ cycles below or above the signal frequency f_s may intercept sufficient energy to produce a detection.

f_s = signal center frequency in cycles per second.
b_s = signal spectral bandwidth in radians per second.
t_d = signal duration in seconds.
T_p = signal pulse-repetition period in seconds.
F_p = signal pulse-repetition frequency in cycles per second.
f_2, f_1 = limits of panoramic receiver sweep band in cycles per second.
D = sweep bandwidth of receiver in cycles per second.
s = receiver sweep rate in radians per second per second.
a = acceptance bandwidth of receiver in cycles per second.
b = receiver acceptance bandwidth in radians per second.
T_s = receiver sweep period in seconds.
θ_{ta} = transmitting-antenna beamwidth in degrees.
t_{ta} = transmitting-antenna look period in seconds.

T_{ts} = transmitting-antenna rotation period in seconds.

θ_{ts} = transmitting-antenna beamwidth in degrees.

t_{rs} = receiving-antenna look period in seconds.

T_{rs} = receiving-antenna rotation period in seconds.

t_{rd} = random delay in seconds.

F_c = frequency of coincidence in cycles per second.

F_{ts} = frequency of revolution of radar antenna in cycles per second.

T_c = coincidence period in seconds.

SRF = scanning-repetition frequency.

The information in Figures 6-8 and 6-9 can be represented in another form which is illustrated in Figure 6-10. Figure 6-10(a) represents the panoramic receiver searching for the pulse train illustrated in Figure 6-8.* Two pulse trains are shown: (1) the upper pulse train is unity whenever a pulse occurs, and zero otherwise; and (2) the lower pulse train is unity whenever the receiver acceptance bandwidth lies within the frequency interval $f_s \pm b_1/4\pi$ occupied by the pulse signal spectrum. Similarly, the two pulse trains in Figure 6-10(b) represent the antenna rotation problem of Figure 6-9. The upper pulse train is unity whenever the receiver site lies within the transmitting beamwidth, and zero otherwise; and the lower train is unity whenever the transmitter site is within the receiving-antenna bandwidth. In both Figures 6-10(a) and (b), a coincidence will be achieved when there is an overlap between the two pulse trains. This occurs at A in Figure 6-10(a). The problem of calculating the probability of coincidence in the case of regular scanning patterns such as those illustrated in Figures 6-8 and 6-9 is thus reduced to calculating the probability of coincidence between two periodic pulse trains. This was recognized by P. I. Richards during World War II and led to his careful study of this problem (reference 11).

FIGURE 6-10 Pulse-train representations of coincidence problems. (a) Sweeping receiver and a pulse signal. (b) Rotating transmitting and receiving antennas.

6.2.3 Unit Coincidence Probabilities

The probability of a coincidence of the receiver acceptance band with a

*With a different time scale.

signal pulse on a single sweep has been called the "unit coincidence probability" (References 13 and 15). As long as the pulse-repetition period is greater than the sweep period, the probability of frequency coincidence alone is the ratio of the duration of the pulse plus the duration of the receiver acceptance time divided by the sweep period

$$P_s{}^{(1)} = \frac{t_d + [(b_s + b)/s]}{T_s} \tag{6-22}$$

where $T_p > T_s$ and $P_s{}^{(1)} = $ the probability of a signal-frequency coincidence on a single receiver scan. For $T_p < T_s$ one has the possibility of coincidence with more than one pulse. When the pulse length t_d is much less than the pulse repetition period T_p, the number of pulses which occur in one sweep period T_s is just

$$n = \text{the next integer} > (T_s/T_p). \tag{6-23}$$

The probability of a frequency coincidence during one scan is then multiplied by this factor and becomes

$$P_s{}^{(1)} = n \left\{ \frac{t_d + [(b_s + b/s]}{T_s} \right\} \tag{6-24}$$

which is valid until $T_p = t_d + [(b_s + b)/s]$. For shorter T_p at least one coincidence is certain to occur and one has

$$P_s{}^{(1)} = 1; \ T_p < t_d + [(b_s + b)/s]. \tag{6-25}$$

For an interception to occur it is necessary that there be coincidence between the receiving and transmitting antenna scans as well as between the frequency of the signal and the receiver frequency. Thus, if the two scans are independent,

$$P^{(1)} = P_a{}^{(1)} P_s{}^{(1)} \tag{6-26}$$

where $P_s{}^{(1)} = $ the probability of an intercept coincidence on a single frequency scan and $P_a{}^{(1)} = $ the probability of an antenna coincidence during a single frequency scan period.

If both antennas are scanning, $P_a{}^{(1)}$ is not easy to calculate because of possible synchronization effects, which are also encountered when one attempts to calculate $P_s(t)$, the probability of at least one signal-frequency

coincidence by a time t. (Recall $P_s^{(1)}$ is probability of coincidence during one scan.) On the other hand, if the receiving antenna is omnidirectional, $P_s^{(1)}$ is just the duty factor of the upper pulse train in Figure 10(b),

$$P_s^{(1)} = t_{ts}/T_{ts}. \qquad (6\text{-}27)$$

Thus, in this case the probability of an intercept coincidence on a single frequency scan becomes

$$p^{(1)} = \frac{t_{ts}}{T_{ts}} \cdot \frac{t_s + [(b_s + b)/s]}{T_s} \qquad (6\text{-}28)$$

(the receiving antenna is omnidirectional and $T_p > T_s$). For very slowly scanning receivers this condition does not hold and the probability of coincidence (Eq. 6-28) increases up to unity when the duration of the coincidence between the receiver acceptance bandpass and the signal spectrum exceeds the duration of one transmitting antenna rotation (Reference 15).

The probabilities of coincidence calculated here, as examples, assume that no overlap is required between the pulse trains of Figure 6-10. Such calculations can be modified to include a minimum duration of coincidence between the pulse trains without difficulty (Reference 15). Equivalently, one can take into account the need for a finite duration of coincidence by one's choice of the signal spectral bandwidth b_s and the receiver acceptance bandwidth b.

6.2.4 Time-Dependent Probabilities

When one attempts to calculate the probability of coincidence as a function of the duration of the searching time, one encounters various problems related to the periodic nature of the events under consideration. Because of this periodicity the mathematical techniques for combinatorial analysis of random events are not generally applicable. It has been suggested, however, that a random delay be intentionally introduced into the receiver sweep pattern to overcome this difficulty (Reference 15). If $T_s \gg T_p$, a random delay t_{rd} can be programmed between each frequency sweep with an average delay approximately equal to T_p. This will render successive sweeps random without introducing an excessive "dead" time. The average dead-time ratio will be just $T_p/(T_s + T_p)$. For $T_s < T_p$, this becomes important. In this case, one can still achieve the effect of randomness by introducing a random delay t_{rd} such that $t_{rd} \approx T_p$ every kth sweep. k should be chosen the next integer above T_p/T_s.

When the successive trials are independent, then the probability of at least one coincidence in m trials (sweeps) is just

$$p^{(m)} = 1 - [1 - p^{(1)}]^m \qquad (6\text{-}29)$$

One notes that $p^{(m)}$ can never become unity. A question of great importance in a search system is the time required before the probability $p^{(m)}$ reaches some specified value. As the time for each trial is $T_s + t_{rd}$, the number of trials, m, required can be obtained from Eq. (6-29) by transposing and taking the logarithm of both sides, and is found to be

$$m = \frac{\ln\left[1 - p^{(m)}\right]}{\ln\left[1 - p^{(1)}\right]} \qquad (6\text{-}30)$$

The time required is then

$$t_{req} = m(T_s + t_{rd}) = (T_s + t_{rd}) \frac{\ln\left[1 - p^{(m)}\right]}{\ln\left[1 - p^{(1)}\right]} \qquad (6\text{-}31)$$

where t_{req} = time required to achieve a probability $p^{(m)}$ of coincidence with
        at least one pulse
   T_s = receiver scan period
   t_{rd} = average receiver delay time between scans
   $p^{(1)}$ = probability of coincidence on a single scan

When successive trials are not independent, the problem of synchronization becomes important. Clearly if T_s/T_p is an integer in Figure 6-10(a), depending upon the relative phasing of the two pulse trains, one can have a signal coincidence on every trial or on no trials. If n is a noninteger but a rational number, one can still have synchronization; but the extremes of the coincidence probability will not be as great. Several authors have studied these effects in some detail (References 9, 10, 15). Matthews considers the long-time average of the fraction of the pulses which coincide with the receiver passband during the transmitting antenna "look" assuming an omnidirectional receiving antenna. By choosing a sweep speed such that the sweep period is less than the look time ($T_s < t_{la}$) and the receiver look time is greater than the signal period $[(b_s + b)/s > T_p]$, one can avoid the synchronization effects and guarantee that the fraction of the pulses intercepted from each transmitting-antenna look is the intercept ratio

$$R = \frac{b_s + b}{2\pi(f_2 - f_1)}. \qquad (6\text{-}32)$$

Under the assumptions $T_s < t_{la}$ and $T_p < [(b_s + b)/s]$, one is bound to intercept at least one pulse on every look of the radar antenna. The average time required to pick up at least one pulse with a probability $P(t)$ is

$$t_{req} = (T_{ts} + T_s)P(t) \tag{6-33}$$

For $P(t) = 1$, this is just the scan time of the antenna plus the scan time of the receiver.

When one cannot achieve satisfactory performance under the assumptions given above, one is forced to face the problem of estimating the probability of coincidence between two periodic wave trains as indicated in Figure 6-10.

FIGURE 6-11   Two periodic wave trains.

This problem has been treated rigorously by Richards (Reference 11). Consider the two wave trains shown in Figure 6-11. Richards shows that if t_1/T_1 and t_2/T_2 are both small and $T_1 \geqq T_2$, the probability of $P(t)$ exceeding $P_0 + Q$ is

$$P^*(t) = 1 - 1.216Q \left\{ 1 + \left[\frac{t(t_1 + t_2)}{T_1 T_2 Q} - 1 \right] \ln \left[1 - \frac{T_1 T_2 Q}{t(t_1 t_2)} \right] \right\} \tag{6-34}$$

where $P_0 = t_1 t_2 / T_1 T_2 =$ probability of coincidence at $t = 0$

$P^*(t) =$ probability that, after a trial of duration t, the probability of intercepting at least one pulse is greater than $P_0 + Q$

This expression is only valid for $(P_0 + Q) < \frac{1}{2}$. It is important to bear in mind that $P^*(t)$ is not the probability of coincidence $P(t)$, but only *the probability* that, by the time t, $P(t) > (P_0 + Q)$. If the time is limited by

$$t \leqq \max (T_1, T_2) \tag{6-35}$$

(max (x, y) means x or y, whichever is greater), then the probability $P(t)$ is given exactly by

$$P(t) = P_0 + \frac{(t_1 + t_2)}{T_1 T_2} t \tag{6-36}$$

W. W. Peterson has been able to extend this result to show that the mean value of $P(t)$ obtained by averaging over all possible ratios of T_1/T_2 and assuming that this ratio is uniformly distributed over all values is

$$\overline{P(t)} \approx P_0 + \left(\frac{t_1 + t_2}{T_1 T_2} \right) t - 0.304 \left(\frac{t_1 + t_2}{T_1 T_2} \right)^2 t^2 \tag{6-37}$$

provided $P(t) \leqq 0.5$ and $t \leqq [(T_1 + t_2)/2T_1T_2]$ (Reference 16).* This quadratic equation can be solved to give the time needed to achieve an average $P(t)$:

$$t = \frac{T_1 T_2}{0.608(t_1 + t_2)} \left\{ 1 - \sqrt{1 - 1.216[P(t) - P_0]} \right\} \quad (6\text{-}38)$$

Equations (6-34) to (6-38) are all based on the requirement of zero overlap between the two pulse trains. If the pulse trains overlap for a time t_0 before a coincidence is scored, the quantity $(t_1 + t_2)/T_1T_2$ in these equations should be replaced by $(t_1 + t_2 - 2t_0)/T_1T_2$.

6.8 Summary of Preceding Sections

In the preceding sections of this chapter some of the concepts and difficulties of the calculation of probability of intercept have been introduced. Signal-detection theory provides one with an estimate of the very best that can be achieved in this direction, and it shows clearly the price one must pay for not knowing what one is looking for. Because of the complexities of receivers which optimize the probability of intercept for a wide range of possible signals, the designer is forced into the use of nonoptimum scanning receivers when his knowledge of the expected signal parameters is poor. This is always likely to be the situation when one designs receivers for COMINT or ELINT applications, for example. On the other hand an intercept receiver for use with a particular jammer or a particular weapon may represent a very different problem. Such a receiver need only look for signals of interest to the particular jammer or weapon involved; and the number of these may be rather well defined. In such a situation the designer may be able to take advantage of the limited number of possible signals to design a receiver capable of much better performance than the general purpose scanning receiver.

In looking into the performance of scanning receiver systems one finds the situation rather less clearly defined than one would wish, particularly if the scanning processes involved are periodic. Nevertheless one can note some general things about scanning receiver systems. In Figure 6-12 are plotted the number of scans necessary to achieve a specified probability of coincidence versus the probability of a coincidence on a single scan. Both the random case given by Equation (6-31) and the Peterson average of the periodic case are plotted. Two things are apparent: (1) the average time for the periodic case does not differ markedly from the time for the random sweep case; and (2)

*In order to obtain manageable expressions for $P(t)$, which must be averaged, Peterson restricts himself to Richard's "open" and "single-overlap" cases. As long as $P(t)$ is less than one-half, these are the only cases possible.
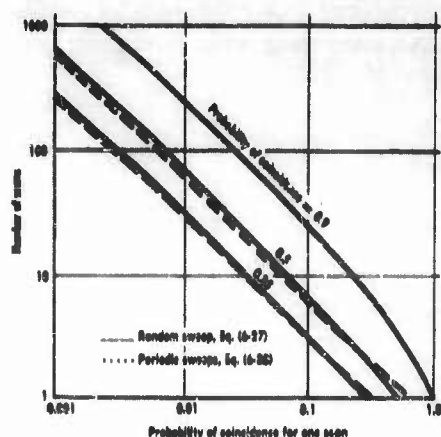
Figure 6-12 Number of scans to achieve a specified probability of coincidence versus probability of coincidence on a single scan.

in both cases over the range of probabilities usually of interest (probability of coincidence per scan less than 0.1), *the time required to achieve a specified intercept probability varies inversely with the probability of intercept per scan.* This is certainly the most important single price that one must pay in going from a wide-open to a scanning system. In some circumstances one can afford the extra time required; in others it is simply not tolerable. To reduce search time one must either scan more rapidly or increase the probability of an intercept on each scan. Intercept systems which involve scanning more than one parameter (frequency and antenna direction, for example) almost always lead to a very low probability of intercept per scan (the slower scan) and hence to very long search times. As a result most practical systems have been limited to the scanning of a single parameter (usually frequency).

When one is looking for signals of short duration, one is naturally led to increased scanning speeds. In the definitions of Section 6.2.2, acceptance bandwidth is specified as though it were independent of the sweep speed. As the sweep speed is increased, one finds that the effective acceptance bandwidth of the receiver increases, and the amplitude of the receiver output decreases. This problem has been studied by several authors (References 17, 18, and 19). As one scans faster and faster, the effective acceptance bandwidth ultimately approaches the entire frequency interval being scanned; and the panoramic receiver becomes a wide-open, broadband receiver! One can show that the output signal-to-noise ratio for such a receiver is the same as for a conventional broadband receiver if optimum video filtering is used (Reference 20). Subsequent studies of the statistical performance of a rapidly scanned panoramic receiver have indicated that from the signal detectability point of view there is never a finite, optimum video bandwidth. Depending upon the parameters of the situation the optimum video filter is always found to be very narrow (an integrator) or very wide (no filter).

6.4 Coincidence Properties of Various Types of Receivers

During the early days of the countermeasures program, the only receivers

available for purposes of interception and identification of frequency were of the "slow-scan" type adapted to be tuned manually across a frequency band in a matter of minutes. Under these conditions, most of the early work on intercept probability was concerned with the determination of optimum scanning-repetition frequencies as a function of radar revolution rates (References 21, 22, 23, and 24). Since the scanning-repetition frequencies (SRF's) of such receivers may be comparable with the frequencies of revolution of searching radar sets, they were not well adapted for use as intercept receivers.

In more recent years, the advent of "rapid-scan" receivers capable of scanning over the entire frequency band in a small fraction of a second has made it possible to examine new scanning-frequency regions, in which the interception ratio is a function principally of the pulse-repetition frequency or of the look period. It has been found that satisfactory interception can be achieved using these higher (SRF's), thereby rendering further examination of the longer-scan region unnecessary.

6.4.1 Slow-Scan Receivers

6.4.1.1 Displacement of the Acceptance Band by Its Own Width in One Revolution Period. The intercept-probability problem can best be approached by examining the manner in which the ratio of pulses intercepted varies as a function of the scanning-repetition frequency, the radar revolution frequency, and the look period or the radar prf. Initially, consider the region in which the receiver scanning-frequency is very small in comparison with the radar-antenna revolution frequency. Specifically, assume that the receiver is tuned at a rate such that its frequency changes by α, its acceptance bandwidth, in megacycles per second, in the period of one radar revolution. Such a condition is indicated diagrammatically in Figure 6-13.
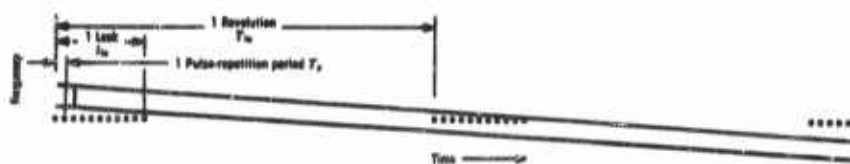
FIGURE 6-13   Slow scan; receiver tunes through own bandwidth in one period of revolution of transmitting antenna.

We now have the following relationships: the ratio of pulses intercepted, or "intercept ratio" is

$$R = \frac{F_0}{F_{tu}} \tag{6-39}$$

where F_o is the frequency of coincidence between the receiver frequency and the radar frequency, and F_{ta} is the frequency of revolution of the radar. One train of pulses is directed at the receiver during each revolution of the radar, and F_{ta} trains of pulses are therefore available for interception during each unit of time. F_o is given by

$$F_o = 1/T_o \qquad (6\text{-}40)$$

where T_o is the time interval between coincidences. In this case, one pulse train or one "look" is intercepted during each scanning period and we have;

$$T_o = T_s \qquad (6\text{-}41)$$

where T_s is the scanning period of the receiver. In turn, T_s is given by

$$T_s = \frac{D}{\alpha} T_{ta} \qquad (6\text{-}42)$$

since the receiver tunes over a band α wide in time T_{ta}, and the scanning bandwidth is D cycles per second.

Substituting Eq. (6-42) and (6-41) in (6-40), we have

$$F_o = 1 \Big/ \frac{D}{\alpha} T_{ta} = F_{ta}\alpha/D \qquad (6\text{-}43)$$

and substitution of this expression in Eq. (6-39) yields the required result,

$$R = \frac{(F_{ta}\alpha/D)}{F_{ta}} = \frac{\alpha}{D}. \qquad (6\text{-}44)$$

Thus, in this case, the ratio of pulses intercepted is equal to the ratio of the acceptance bandwidth of the receiver to the scanning bandwidth of the receiver.

The maximum time required to intercept the signal is, evidently, the coincidence period

$$T_o = \frac{D}{\alpha} T_{ta} \qquad (6\text{-}45)$$

By way of example, assume that we have the following typical case:

$$T_{ta} = 30 \text{ seconds}$$
$$D = 2000 \text{ Mcs}$$
$$\alpha = 200 \text{ Mcs}$$

Then the ratio of pulses intercepted is found, from Eq. (6-44), to be

$$R = 200/2000 = 0.1 \qquad (6\text{-}46)$$

and Eq. (6-45) shows that the maximum time required to achieve interception is

$$T_o = (2000/200) \times 30 = 300 \text{ seconds} = 5 \text{ minutes.} \qquad (6\text{-}47)$$

Although the time required for interception is long, the relationship given by Eq. (6-4·) and (6-45) are useful, and this region of operation is of practical importance because it is at least certain that an interception will be obtained during each scanning period of the receiver. The assumption is made, of course, that the acceptance band of the receiver is sufficiently wide so that the receiver remains on at the radar frequency during the entire period of each intercepted look. This condition actually exists, even in the case of narrowband receivers.

In the above examples, a 10 percent acceptance-band to scanning-band ratio has purposely been assumed because this ratio applies to rapid-scan receivers, which will form the basis of the discussion of other operating regions in which the scanning-repetition frequency is higher. It is readily seen from Eq. (6-45) that, if a conventional narrowband ($a = 2\text{Mcs}$) superheterodyne intercept receiver is used in this manner, scanning over the same band, D, the maximum time required for interception may be 100 times as great, or 500 minutes ($8\frac{1}{3}$) hours. This points up one of the difficulties of using such receivers as intercept receivers.

6.4.1.2 Scan Period Comparable with Revolution Period. If it is now assumed that the SRF is raised so that the acceptance band of the receiver is displaced more than its own width in one revolution period, we enter a region in which little can be predicted regarding intercept effectiveness. The region discussed in Section 6.4.1.1 above involves SRF's of the order of 10 scans per hour. We now consider SRF's ranging from this figure through the region in which the SRF is comparable with the radar revolution frequency, say 100 revolutions per hour, to the region in which the SRF assumes a value, say 20 scans per second, comparable with the reciprocal of the time duration of a radar look. This is a very extensive part of the SRF spectrum, including as it does, SRF's from the order of magnitude of 10^{-8} per second to the order of magnitude of 10 per second.

As soon as the acceptance band of the receiver is displaced more than a in the period of one radar revolution, it can be seen that there is a possibility of the pulses of a given radar look occurring at an instant when the receiver

is not on the radar frequency. Under these conditions no interception occurs. As the SRF is still further increased, a series of values is reached at which the receiver "gate" is open at the radar frequency every nth look. For example, if the SRF is one-quarter the radar revolution frequency, and the phase relations are favorable, the receiver will intercept every fourth look. On the other hand, if the phase relations are not favorable under these conditions, the receiver will fail to intercept any looks at all.

Further increase in the SRF brings it to the value where it is equal to the radar revolution frequency. At this value of SRF, every look will be intercepted if the phase relations are favorable, but a complete failure to intercept any pulses will occur if the phase relations are not favorable.

When the SRF is greater than the radar revolution frequency, a situation prevails which is similar to that which exists when the SRF is less than the radar revolution frequency. "Beats" occur between the receiver and radar gates, and the ratio of pulses intercepted may be large or small depending on the beat frequency and the phase relationships involved.

This case is most easily analyzed by considering the condition which exists in the region where the SRF is closely comparable to the radar revolution frequency. Figures 6-14, 6-15, and 6-16 are diagrammatic representations of this region. (In Figures 6-14, 6-15, 6-16, 6-17, and 6-18 only about every fifth pulse is shown because of drafting limitations.) In each of these figures, the top line represents the times of occurrence of the radar looks, the center of each of which is indicated by the short vertical lines. The bottom line represents the times at which the center of the receiver gate is at the radar frequency. In what follows, no attempt is made to appraise the effect of the relationship between the acceptance bandwidth and the duration of the radar look. The results obtained herein might be expanded to take into account the fact that these occurrences actually have finite duration; but the value of such effort would seem to be problematical, since it is quite evident that this is not an optimum operating region.

First consider Figure 6-14. This is the case where the SRF is exactly equal to the radar revolution frequency, and the phase relations are favorable. It is quite evident without any analysis whatsoever that under these conditions all looks are intercepted, R is unity, and T_e is T_{fa}, if we assume a coincidence at $t = 0$.

However, it is also evident that, in order to maintain this perfect intercept effectiveness, it is necessary that the gates be open at *exactly* the same time. Since they are shown at being infinitely narrow, even the slightest time lag or gain will result in losing the interception entirely. The extreme case of this type is represented by Figure 6-15, in which the radar is exactly half a cycle out of phase with the receiver, and no pulses are intercepted.
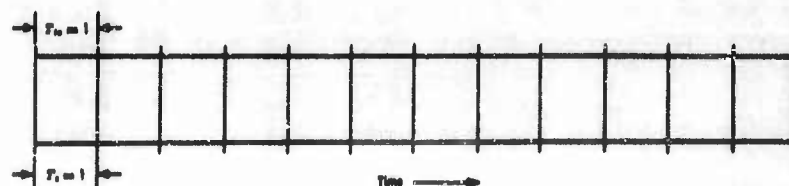
FIGURE 6-14   Slow scan; scanning repetition period equal to, synchronous with, and in phase with revolution period.
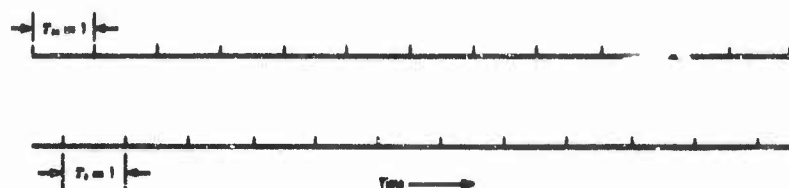
FIGURE 6-15   Slow scan; scanning repetition period equal to, synchronous with, and out of phase with revolution period.
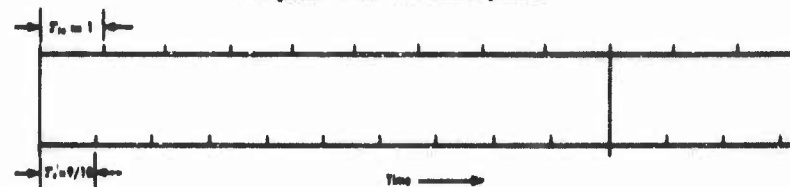
FIGURE 6-16   Slow scan; scanning repetition period comparable with but shorter than revolution period.
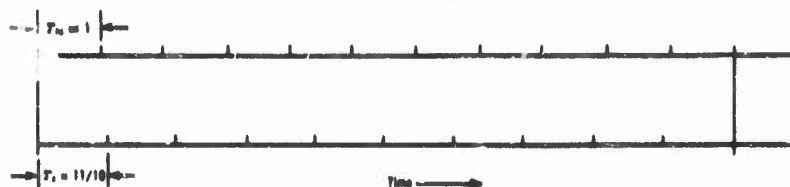
FIGURE 6-17   Slow scan; scanning period comparable with but longer than revolution period.
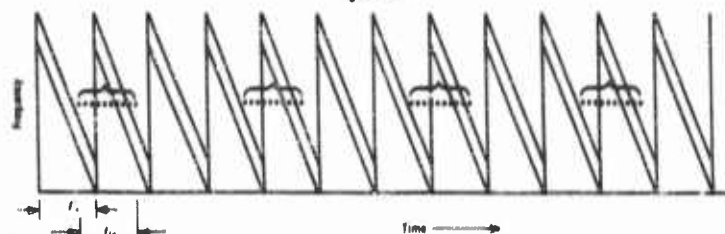
FIGURE 6-18   Rapid scan; scanning repetition period comparable with radar look.

The conclusion is, of course, that when the SRF is *exactly* equal to the radar revolution frequency, the intercept ratio R may be either 1.0 or 0, depending upon the phase relations. In the case of practical receivers in which the finite width of the gates permits the interception of the trailing edge or of the leading edge of a pulse train, this conclusion would be modified to state that when the SRF is *exactly* equal to the radar revolution frequency the intercept ratio R may lie between 1.0 and 0. This would be a trivial result, except for the fact that in certain other regions R is much more narrowly limited, regardless of phase.

Because it is practically impossible to maintain two periodic functions exactly in step over any appreciable period of time, neither Figure 6-14 nor Figure 6-15 has much practical significance. Instead, the conditions demonstrated by Figures 6-16 and 6-17 are those which will prevail in this region most of the time. Figure 6-16 represents the case where the SRF is slightly greater than the radar revolution frequency or, in other words, where the scanning period is slightly shorter than the radar revolution period.

In analyzing Figure 6-16, we observe that, at the end of the first revolution of the radar antenna the receiver gate is $T_{is}/10$ ahead. At the end of the second revolution, the receiver gate is $2T_{is}/10$ ahead, and it increases its lead by $T'_{is}/10$ each revolution, until, after 10 revolutions, it is ahead by T_{is}; and a coincidence occurs on the ninth radar look.

This would lead to the conclusion that a coincidence will occur whenever enough revolutions have been passed to advance the scanning gate by one revolution period. This conclusion is approximately correct, but note needs to be taken of the fact that exact multiples of T_s and T_{is} may be involved and that the advance may be more than one revolution period. If, in fact, T_s were $2T_{is}/10$, the advance per period would be $2T_{is}/10$ and the gate would advance four revolution periods in order to produce a coincidence. Also, the advance must be an *integral* number of revolution periods, since the gates are shown as being infinitely narrow.

A similar situation prevails when $T_s > T_{is}$, as shown in Figure 6-17.

Extreme variations in R and T_o occur in this region with very small variations in T_s/T_{is}, and since there is no possible way in practice to fix the value of T_s in relation to T_{is}, or even to predict it, it would be impossible to utilize any optimum relationship which might be established. It is concluded that, since there are other regions in which optimum values of the operating parameters can be easily determined and easily achieved in practice, exact analysis to determine R and T_c in this region in which SRF's are comparable with radar revolution frequencies would not be justified, at least for the purpose of this discussion.

6.4.2 Rapid-Scan Receivers

6.4.2.1 Scan Period Comparable with Duration of Radar Look.
As the SRF is now further increased, the scanning period begins to approach
the value where it is comparable in length to the duration of a radar look.
Typically, the length of a radar look may be in the vicinity of $\frac{1}{20}$ second. In
this section therefore, SRF's in the neighborhood of 20 cycles per second are
discussed. The pulse length t_d and the spectral duration b_s/s, are disregarded,
because they are both small in comparison with b/s, the time the receiver
gate is open on the frequency of the pulse.

This is one of the optimum regions, in which performance can be predicted
with satisfactory accuracy. The results of Eq. (6-28) do not apply since the
antenna scanning process is not random. The relationships developed in this
section, though simple, will therefore be found to be significant from a prac-
tical standpoint. Considering, first, the case when the scanning period T_s is
exactly equal to the duration of a radar look t_{ls} as shown in Figure 6-18, it
is evident that some pulses will be intercepted during each scanning period.

The number of pulses intercepted out of any given look will be the product
of the pulse-repetition frequency F_p and the length of time the receiver gate
is open on the radar frequency. The intercept ratio (the fraction of pulses
intercepted from a given radar look) will then be this quantity divided by
the number of pulses in a look. Since the receiver gate is open on the radar
frequency for a time duration $\alpha T_s/D$, and the number of pulses in a look is
$t_{ls}F_p$, we have, for $T_s = t_{ls}$,

$$R = \frac{F_p \alpha T_s/D}{t_{ls}F_p} = \frac{\alpha}{D} \cdot \frac{T_s}{t_{ls}} = \frac{\alpha}{D} \tag{6-48}$$

It is also evident that the maximum length of time required to produce an
intercept, if the look is already there when the scanning cycle starts, is T_s.
However, if time is reckoned from the moment when the receiver is turned
on, there may be no pulse train available for interception until the radar
antenna comes around and looks at the receiver, d if this definition is
used the maximum intercept time will be T_{ls}.

As the SRF is further increased, the relationships of this section (averaged
over a number of sweeps) continue to apply until the time the receiver gate
is open becomes less than the period of the prf, under which conditions it is
possible that the gate may exist, at the radar frequency, only between pulses.
This upper limit is given by

$$(\alpha/D)T_{s\,min} = T_p$$

$$T_{s\,min} = (D/\alpha)T_p \tag{6-49}$$

For example, in the case of a rapid-scan receiver with an acceptance band to scanning-band ratio, α/D, of 0.1, searching for a radar operating at a prf of 1000, the minimum scanning period applicable to this region is

$$T_{s\,min} = 10 \times 1/1000 = 0.01 \text{ second} \tag{6-50}$$

6.4.2.2 Scanning-Repetition Frequency Comparable with PRF. We now enter the region in which it is possible for the acceptance band of the receiver to cross the radar frequency between pulses. Considerations applying to this region are similar to those already discussed in Section 6.4.1.2 in connection with the condition when the scanning period is comparable with the revolution period. The discussion of Section 6.4.1.2 will not, therefore, be repeated here.

Where a random delay is employed, the analysis of Section 6.2.4 applies here. In any case, this is a useful operating region, because as shown by Dethlefsen (Reference 10), when T_s and T_p are small, phase and frequency instability produce a desirable smoothing effect in a short period of time and eliminate the sharp peaks and nulls of synchronism. These considerations are discussed in more detail in Section 6.4.3.

6.4.2.3 Scan Period Comparable with Pulse Length. Technical developments have now made it possible to operate with a scanning period as short as the length of the pulse. In this region, it is obvious that perfect intercept effectiveness is achieved. Since the receiver scans over the entire band in the period of a pulse, it is impossible to miss any pulses, and every pulse will be intercepted. Therefore, the intercept ratio is 1.0 and the maximum time required to produce interception is either T_s or T_{is}, depending upon whether time is reckoned from the start of a radar look or from the time when the receiver is turned on.

The usefulness of this region of operation is limited principally by techniques. Since by definition the time the gate is open on the radar frequency is shorter than the pulse length, it follows that the gate cuts the pulse into segments which are only a fraction of a pulse length long. The result is to increase the video bandwidth requirements of the receiver, to reduce the amount of energy available for display purposes, and to increase the writing-speed requirement of the display oscilloscope. It has been shown on theoretical grounds (Reference 25) that these extended design requirements can be met with only a relatively small amount of circuit and component refinement, and receivers have been constructed.

6.4.3 The Effect of Instability

In Section 6.2.4 the effect of introducing a random delay into the receiver sweep pattern is discussed and analyzed. It has been found that such random delays are actually present, even if not intentionally introduced, in view of the phase and frequency instability of both the radar and the receiver. The instability of these parameters is, in fact, so large that, over a relatively short period of time, the peaks and nulls due to synchronization are smoothed out, and an averaged, integrated value is obtained.

A general analysis along these lines has been completed and verified experimentally by Dethlefsen (Reference 10). As in Sections 6.2.3 and 6.2.4, the statistical principle involved is that, if a certain number of events occur at random at a certain average rate over a long period of time, and if the continuum is sampled at random during that same period of time, then the proportion of events detected will be equal to the proportion of time during which the sampling mechanism is receptive. Applied to the intercept problem this means that, if the average number of pulses directed (at random intervals) at the receiver over a long period of time is A and if during that same period of time the receiver is receptive at the radar frequency for a fraction of time represented by B, then the number of pulses intercepted will be AB. More specifically, if the prf is F_p, the receiver acceptance bandwidth is a, and the receiver scanning band is D, then the receiver will be open at the radar frequency for a fraction of time a/D and the rate at which pulses will be intercepted will be $F_p a/D$, regardless of the frequency and phase relations involved, providing they are both random and considering frequency scanning only. Then it follows that, with these same limitations, the intercept ratio is a/D.

The validity of this line of reasoning is quite evident, provided that it can be accepted that the processes involved are actually random. It may seem surprising that this is true in the case of the intercept problem, in view of the care which has previously been taken to determine exact frequencies, bandwidths, phase relationships, etc. However, careful tests have been made using simulators, and it has been determined that the instability encountered in actual practice is, indeed, sufficient so that the statistical result applies with considerable exactness. In fact, attempts to maintain synchronism between T_s and T_{is} or T_p with the objective of operating at one of the peaks or nulls of the synchronous condition always met with failure; after a relatively short period of time, synchronism was lost and the intercept ratio became a/D.

It may therefore be stated, by way of summarizing, that in the practical intercept case the average intercept ratio is always a/D, regardless of the nominal values of T_s, T_p or T_{is}, provided that the observations are taken over a long period of time.

The question still remains as to what is a long period of time, in this case.

Tests have shown that this means that the period of observation must be long in comparison with any of the variables T_s, T_{ts} or T_p. These relationships are therefore not valid in the case of short-burst transmissions.

6.5 Summary of the Coincidence Problem

Taking the considerations of Section 6.4 into account we are now able to draw the following conclusions regarding the relative merits of the various regions of operation which have previously been considered.

(A)　Receiver tunes through α in T_{ts} (Section 6.4.1.1)

　　　This region is established as valid and predictable without relying on statistical considerations. The maximum time of intercept is, however, too long, and this is not a desirable operating region.

(B)　T_s comparable with T_{ts} (Section 6.4.1.2)

　　　Here, statistical analysis indicates that a satisfactory intercept ratio can be obtained if the observations are continued over a sufficiently long period of time. However, this means that the period of observation must be long in comparison with the radar revolution period T_{ts}, which is in itself long. Thus, satisfactory intercept effectiveness cannot be expected in a period of time less than a half-hour, and this is not a desirable operating region.

(C)　T_s comparable with t_{ts} (Section 6.4.2.1)

　　　This is the most satisfactory of all the regions discussed. Performance is predictable without resorting to statistical considerations. The intercept ratio is always α/D and the maximum interception time not greater than T_{ts}.

(D)　T_s comparable with T_p (Sections 6.2.3, 6.2.4, and 6.4.2.2)

　　　Although the complications of synchronism occur here, statistical considerations show that this is a desirable operating region, because T_s and T_p are both small, and the time of observation therefore can be relatively small and still remain large with respect to T_s and T_p, producing an intercept ratio of α/D.

(E)　T_s comparable with t_d (Section 6.4.2.3)

　　　Perfect intercept effectiveness is attained in this region, and each pulse is intercepted individually.

REFERENCES

1. Woodward, P. M., "Probability and Information Theory, with Applications to Radar," McGraw-Hill, New York, 1955.

2. Peterson, W. W., T. G. Birdsall, and W. C. Fox, "The Theory of Signal Detectability," *Trans. IRE*, PGIT-4, September 1954.

3. Reich, E., and P. Swerling, "The Detection of a Sine Wave in Gaussian Noise," *J. Appl. Phys.*, Vol. 24, p. 289, 1953.

4. Middleton, D., and D. Van Meter, "Detection and Extraction of Signals in Noise from the Viewpoint of Statistical Decision Theory," *Soc. for Ind. and Appl. Math.*, Vol. 4, No. 4, December 1955 and Vol. 5, No. 1, March 1956.

5. Peterson, W. W., and T. G. Birdsall, "The Theory of Signal Detectability," Tech. Report, No. 13, (DA-36-030 SC-15358), EDG, The University of Michigan, Ann Arbor, June 1953.

6. Birdsall, T. G., and W. W. Peterson, "Probability of Correct Decision In a Forced Choice Among M Alternative," Quarterly Progress Report No. 10, EDG, The University of Michigan, Ann Arbor, April 1954.

7. Green, D. M., and T. G. Birdsall, "The Effect of Vocabulary Size on Articulation Score," Tech. Report No. 81, EDG, The University of Michigan, Ann Arbor, January 1958.

8. Hollywood, J. M., "Probability of Intercepting Radar Signals by Search Receivers," Report No. R-2843, ASTIA Code AD-166, NRL, Washington, D. C., June 1946. (CONFIDENTIAL)

9. Matthews, A. R., "An Analysis of Desirable Characteristics of Microwave Intercept Receivers," Tech. Report No. 4 '(N6 onr 25132), ERL, Stanford University, Stanford, California, June 1952.

10. Dethlefsen, D. G., "An Experimental Analysis of Signal Interception by Rapid-Beam Microwave Receivers," Tech. Report No. 13 (N6 onr 25132), ERL, Stanford University, Stanford, California, June 1952.

11. Richards, P. I., "Probability of Coincidence for Two Periodically Recurring Events," *Annals of Mathematical Statistics*, Vol. XIX, No. 1, pp. 16-29, March 1948.

12. Fessler, A. L., et al, "Probability of Intercept," Studies of Operational Aspects of Electronic Intercept and Analysis, Report No. 44-F, Vol. 2, pp. 98-184, Haller, Raymond and Brown, Inc., State College, Pennsylvania, April 1954. (SECRET)

13. Kiel, A., "On The Interception of Recurrent Signals," Tech. Report No. AF-41 (AF 33(616)-3374), Rad. Lab., Johns Hopkins University, Baltimore, Maryland, October 1957. (SECRET)

14. Peterson, W. W., "Application of Statistical Techniques To Search and Intercept," Quarterly Progress Report No. 6, Task Order EDG-3 (DA-36-039 SC-15358), The University of Michigan, Ann Arbor, March 1953. (SECRET)

15. Enslow, Jr., P. H., "Some Techniques for the Analysis of Intercept Probability in Reconnaissance Receivers," Tech. Report No. 516-1 (DA-36-039 SC-78296), Systems Techniques Laboratory, Stanford University, Stanford, California, June 1959.

16. Quarterly Progress Report No. 6; Task Order EDG-3; "Application of Statistical Techniques to Search and Intercept", Dec. 1, 1952, to Feb. 28, 1953, pp. 13-15, Electronic Defense Group, The University of Michigan, Ann Arbor, Mich., (DA-36-039 SC-15358).

17. Batten, H. W., R. A. Jorgensen, A. B. Macnee, and W. W. Peterson, "The Response of a Panoramic Receiver to CW and Pulse Signals," *Proc. IRE*, Vol. 42, pp. 948-956, June 1954.

18. Grigsby, J. L., "An Investigation of the Sensitivity of Scan Receivers Having Large RF Bandwidth," Tech. Report No. 553-1, SEL, Stanford University, Stanford, California, November 1958.

19. Peterson, W. W., and T. G. Birdsall, "Signal Detection with a Panoramic Receiver," Tech. Report No. 38, (DA 36-039 SC-63203, EDG, The University of Michigan, Ann Arbor, June 1955.

20. Wilson, Q. C., III, and D. W. Fife, "Statistical Measurements of the Detection of a Continuous Signal by a Panoramic Receiver," Tech. Report No. 105, EDG, The University of Michigan, Ann Arbor, May 1960.

21. Foster, D., "The Use of Continuously Rotating Direction Finders Against Signals of Varying Intensity," Report No. 411-158, February 17, 1945, Radio Research Laboratory.

22. Richards, P. I., "Probability Formulas for Simultaneous Periodically Recurring Events," Report No. 411-171, May 5, 1945, Radio Research Laboratory.

23. Foster, D., "Optimum Directivity Patterns for Search Antennas," Report No. 411-192, May 23, 1945, Radio Research Laboratory.

24. Hulstede, G. E., "The Effect of Signal Intercept Probabilities on Search-Receiver Design," Report No. 411-272, October 10, 1945, Radio Research Laboratory.

25. Harris, D. B., and C. B. Crumly, "Display Considerations in the Rapid-Scan Receiver," Report No. 3 (Task 33), February 15, 1952, Electronics Research Laboratory.

7

Detection and Analysis of Signals

T. G. BIRDSALL, H. BLASBALG

7.1 Problems of Signal Detestability

7.1.1 Introduction

The problem of detection of signals in the presence of random noise and interference is essentially a problem of deciding between two possibilities: signal was present in the noise and interference, or, only noise and interference was present. The problem is statistical in nature, because the noise is a random process, and so the resulting theory of signal detectability is a specific application of the more general statistical decision theory. As a theory for receiver design it differs from the usual signal-to-noise ratio approach to design in that detectability theory designs by considering first the objectives of the receiver and works back toward the design of the circuits, contrasted to beginning with the basic circuit design and optimizing parameter values to achieve optimum signal-to-noise ratio. Because of this change of approach, designs based on the detectability theory can claim to be truly optimum. Of equal importance is the realization that practical, non-optimum, receivers can be rated against the optimum, so that one knows how much could be gained by further improvement.

It was in connection with the development of radar that the modern theory of signal detection had its beginning. North's paper is a notable early work (Reference 1). North uses the ideas of probability of detection and false alarm probability in defining optimum detection, and he discusses the optimization of i-f filter transfer function, detector characteristic, and integration

for a radar system. Siegert was the first to propose the use of the statistical theory concerning hypothesis testing to solve the problem of optimum detection of threshold radar signals. A brief account of this "Theory of the Ideal Observer" appears in *Threshold Signals* by Lawson and Uhlenbeck (Reference 2).

The theory of target detection by pulsed radar received very thorough treatment by Marcum in an unpublished Rand Corporation report (Reference 3). Both optimum and commonly used non-optimum types of systems are considered. Formulas and curves are given for false alarm probability and probability of detection. This report is still a good reference on signal detection in conventional radar. Several other treatments of the radar problem have appeared. The paper by Slattery (Reference 4) treats the problem of detection of a single pulse: dissertations by Hanse and Schwartz (References 5 and 6) treated phases of the radar problem. The radar problem has also been treated by Middleton (References 7 and 8). This treatment features an even greater emphasis on statistical theory. The application of sequential analysis was suggested for the first time in the literature.

Woodward and Davies, also working initially on radar development, made the first steps toward a general theory of signal detection (References 9-12). They introduce the idea of a receiver having *a posteriori* probability (i.e., an estimate of the probability that a signal was sent) as its output rather than a yes-no answer to the question of whether or not a signal was sent. They show that such a receiver gives a maximum amount of information. They treat the case of an arbitrary signal known exactly or known except for carrier phase with no more difficulty than other authors have had with a sine-wave or pulse signal. The noise is assumed to be white and bandlimited.

The most important advance toward a general theory of detection was made by Grenander (Reference 13) in his paper on "Stochastic Processes and Statistical Inference." This paper is mathematical, and makes no mention of noise or the detection of signals. The problems treated, however, are precisely those of detecting signals in noise and estimating signal parameters in the presence of noise. The method for handling arbitrary noise and signals is developed, and a few examples are worked out using gaussian noise.

While the concepts in this theory can be expressed quite simply, complete description of an optimum receiver and evaluation of probability of detection and false alarm probability involves much numerical work in all but the very simplest cases. The work so far reported on efforts toward determination of optimum receivers and their performance falls in three classes according to the manner in which the simplification is made.

In the first case, the noise was assumed gaussian, but with arbitrary spectrum (or autocorrelation). Only two signal ensembles have been studied in detail. Reich and Swerling (Reference 14) studied the case of an amplitude

modulated signal known except for carrier phase. In References 15 and 16 Davis reported on the case of a signal known exactly. Even for these very simple cases the solutions are quite involved.

If secondly, the noise is further restricted to the white bandlimited gaussian noise, the formulas simplify very greatly. This case is studied in detail in Electronic Defense Group Technical Report No. 13 (Reference 17). The report includes a general introduction to the problem of signal detectability, and specific derivations of the optimum receiver and probabilities of detection and false alarm for a number of cases of practical interest. The special cases which are presented were chosen from the simplest problems in signal detection which closely represent practical situations.

A third approach was used by Middleton (References 18 and 19). The method he chose for simplifying the problem was to expand the logarithm of the likelihood ratio in a power series in terms of sample values, and then to keep only first degree terms if they are not zero, or second degree terms, if the first degree terms vanish. Middleton calls the first case the coherent and the second the incoherent case. Then the coherent case is reduced to the study of a linear form and the incoherent case to the study of a quadratic form, both of which, in the mathematical area, have been studied extensively. This approximation involves no error for the case of a signal known exactly and is a good approximation for the case of the signal known except for phase. For other cases, it is difficult to say how good the approximation will be.

The paper by Urkowitz (Reference 20) points up a problem in designing filters for optimum detection. While this paper is concerned with maximizing signal-to-noise ratio, the same problem arises if a statistical approach is used (References 15 and 17).

7.1.2 Fundamental Detection Problem, Theory

The fundamental, or simplest, detection problem involves a decision between two alternatives. The decision is based on the receiver input that occurs during a preselected finite time, and takes into account the parameters of the noise, the various signals that might occur and their relative probabilities of occurrence, and how the noise and signal combine. The receiver (physically, this includes both hardware and/or human being) then processes the receiver input and yields a binary output, for one alternative and against the other. In the usual detection problem the two alternatives are "There was a signal present," and "There was only background noise present at the receiver input." In binary communications such as frequency shift keying (FSK), the two alternatives are "The incoming signal was at the Mark frequency" and "The incoming signal was at the Space frequency."

The problem, called the fundamental problem, has two assumptions that

make it simple: (1) It is assumed that the two alternative decisions correspond to two mutually exclusive and exhaustive classes of causes, and (2) that except for a recalcitrant set whose probability is zero, the probability of occurrence, or the probability density of occurrence, of each possible receiver input is known for each of the two classes of causes. Repeating in a little more depth, the first assumption is that, corresponding to the two possible alternative decisions to be made, there are two possible classes of causes, or simply "causes". This is really the origin of the problem, since it assumes that one is interested in a binary decision because there were indeed two causes to be distinguished from each other. In order to make any sensible decision a given cause (such as a specific signal) must belong to one class or the other, and not jump back and forth, hence "exclusive". The real assumption is the word "exhaustive", which means that no other cause than those listed in the two classes can occur. Thus, if one is trying to detect WWV of 10 megacycles in the presence of noise, with the alternative cause being just noise alone—WWV off the air—then the assumption is that no other transmission is operating on 10 megacycles for that period. The second assumption means that the situation has to be well enough specified and the random parts of the causes be sufficiently stable to allow specific assignment of probabilities. The actual mechanics of determining these probabilities is a major part in the solution of any specific problem.

For the moment assume that one has a specific detection problem for which the above assumptions are satisfied: the only two possible causes of the receiver input are noise alone, signified by N, and signal plus noise, signified by SN; and that for each possible receiver input $x(t)$ the probability densities are known. Specifically, the function $f_N(x(t))$ is the probability density that $x(t)$ will be the receiver input when the cause is noise alone, and the function $f_{SN}(x(t))$ is the probability density that $x(t)$ will be the receiver input when the cause is signal plus noise. The two alternative decisions that the receiver (the decision device) can make are to sound the alarm A that a signal was present, or to conclude that no signal was present, no A or B. When the actual cause is noise alone, these decisions correspond to a false alarm, and a correct no-alarm respectively. The probabilities of these occurrences (conditional to the cause being noise alone) are:

Conditional probability of false alarm $= P_N$ $^{1})$

$$= \int_A^f f_N(x(t))\, dx \qquad (7\text{-}1)$$

Conditional probability of correct-no $= 1 - P_N(A)$

where the integral is taken over all $x(t)$ for which the decision device reaches

the decision A. When the actual cause of the input is signal and noise, the two decisions correspond to a hit and a miss, and the probabilities of these occurrences (conditional to the cause being signal and noise) are:

Conditional probability of detection $= P_{SN}(A)$

$$= \int_A f_{SN}(x(t))\, dx \qquad (\ \cdot 2)$$

Conditional probability of a miss $= 1 - P_{SN}(A)$

The performance of any decision device operating in this specific detection problem can be summarized by one of the numbers in Eq (7-1) and either the hit or miss probability in Eq (7-2).

If these two numbers are plotted against each other on graph paper, the point is called the "receiver operating point", and the total of all such points for all decision devices make up the "receiver operating region". The upper boundary of this region is called the optimum "receiver operating character-
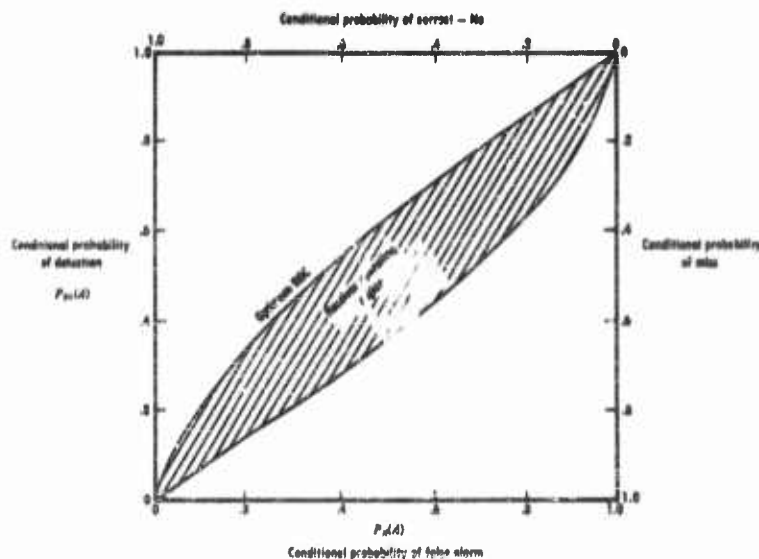
FIGURE 7-1 The Detection ROC

istic" (ROC) for this specific problem. These are sketched in Figure 7-1. Several other characteristics about the detection ROC should be noted before proceeding. The diagonal inside the receiver operating region corresponds to equal conditional probabilities of false alarm and detection, that is the

probability of an alarm is independent of the cause. This is called chance performance, because a decision based on a random selector and independent of the receiver input could achieve this type of performance. The optimum ROC has been shown as a smooth curve, with decreasing (nonincreasing) slope, and the operating region as a simple convex region containing the chance diagonal. Both of these characteristics can be proved to hold in all problems under very general assumptions, but the proof must be omitted here.

For a moment let us concentrate on the upper boundary and why it is called optimum. Choose any operating point in the region that is not on the upper boundary. Then the operating point on the upper boundary directly above it will have the same probabilities when the receiver input is due to noise, but will have a higher probability of detection; in fact, it has the highest probability of detection that can be achieved at the specific probability of false alarm. In the same manner, consider the operating point on the upper boundary that is horizontally to the left of the original operating point. It has the same detection and miss probabilities, but lower false alarm probability and higher correct-no probability. Both of these upper boundary operating points are better (at least, not worse) than the original interior operating point under any definition of "better" where correct responses are advantageous and errors are disadvantageous. Under certain types of operation where bluffing is introduced this may not be true, but if such intentional-mistake operation is ruled out, then any definition of optimum will take on its maximum on this upper boundary. Any receiver which could be adjusted to operate everywhere along this curve would truly be the optimum receiver for this detection situation.

As throughout this discussion of detection theory, no attempt will be made to give proof of the existence of this optimum receiver, but a short intuitive discussion that parallels the proof will be given. If one considers all decision processes, all receivers, that have the same conditional probability $P_N(A)$, one is considering all ways of dividing the receiver inputs $x(t)$ into two sets. The first set will lead to an alarm A and the second set will not, with the restriction that exactly enough receiver inputs $x(t)$ are in the alarm set so that the false alarm probability is the desired value

$$P_N(A) = \int_A f_N(x(t))\, dx = \text{preselected value.} \qquad (7\text{-}3)$$

The corresponding detection probability is

$$P_{SN}(A) = \int_A f_{SN}(x(t))\, dx = \int_A \left[\frac{f_{SN}(x)}{f_N(x)} \right] f_N(x)\, dx \qquad (7\text{-}4)$$

Since the upper boundary has the maximum value of $P_{SN}(A)$ that can be obtained for the preselected value of $P_N(A)$, it must correspond to a selection of alarm set A for which the bracketed function in this last integral is as large as possible. The bracketed function is a function of the receiver input, and is called the "likelihood ratio"

$$l\left(x(t)\right) = \left[\frac{f_{SN}(x(t))}{f_N(x(t))}\right] \tag{7-5}$$

The optimum receiver, the likelihood ratio receiver, processes each input in such a way that the receiver output is the likelihood ratio of the input waveform—one number out for each observed waveform in. The final decision mechanism of the receiver is a comparitor unit that gives an alarm if the likelihood ratio is above a cut level, and no alarm if the likelihood ratio is below the cut level. If the probability of the likelihood ratio being precisely at the cut level is not zero, then an additional mechanism must be introduced to apportion these occurrences between alarms and non-alarms as is operationally desired. For a specific $P_N(A)$ the cut level is adjusted to yield that probability of alarms when the input is caused by noise alone, and the resulting probability of detection will be maximized.

Since the function of the likelihood ratio receiver is that it partitions the input waveforms according to high and low likelihood ratio parts, any device whose output is a monotone function of the likelihood ratio would perform an identical operation if the cut levels are set to corresponding values.

One advantage of the fundamental detection problem is that problems formulated in terms of maximizing some specific quantity, such as the expected monetary return from a decision, separate into two parts. The optimum processing of the receiver input always turns out to be a monotone function of likelihood ratio, because one always does best by operating on the upper boundary of the ROC region. The specific quantity, or criterion, to be maximized merely specifies which point on the upper boundary to use. Some of the common criteria and the corresponding operating points are listed below.

1. The Weighted Combination Criterion: that criterion that maximizes $P_{SN}(A) - wP_N(A)$.

   Solution: w is the likelihood ratio cutoff, say there is signal present if $l(x) \leq w$.

2. The Neyman-Pearson Criterion: That criterion such that (1) $P_N(A) \leq k$ with maximum $P_{SN}(A)$ over all criterion satisfying (1).

   Solution: The likelihood ratio cutoff β set such that $P_N(A) = k$.

3. Seigert's Ideal Observer: That criterion that minimizes total error.
   Solution: Likelihood ratio cutoff $= P(N)/P(SN)$, the ratio of *a priori*
   probabilities.

4. Expected Value Observer: That criterion that maximizes the total expected value, where the individual values are: V_D, the value of detection; V_M, the value of a miss; V_Q, the value of peace and quiet; V_F, the value of a false alarm. (It is assumed that these latter two are less than the former two.)

   Solution: Likelihood ratio cutoff $= \dfrac{P(N)}{P(SN)} \quad \dfrac{V_Q - V_F}{V_D - V_M}$

5. Minimum Risk Criterion: Let $V_F - V_Q = r_N$, the risk when N is the cause, and let $V_M - V_D = r_{SN}$, the risk when SN is the cause. Minimize the average risk, $R = P(SN) \, r_{SN} + P(N) \, r_N$.

   Solution: Likelihood ratio cutoff $= \dfrac{P(N)\sqrt{\tilde{N}}}{P(SN)\sqrt{SN}}$

6. Information Observer: That criterion that maximizes the reduction in Uncertainty (Shannon Sense) as to whether a signal was sent or not.

   Solution: The cutoff β is the solution to the equation

   $$\beta = \frac{P(N)}{P(SN)} \quad \frac{\log P_{B(\beta)}(N) - \log P_{A(\beta)}(N)}{\log P_{A(\beta)}(SN) - \log P_{B(\beta)}(SN)}$$

   where B means not A.

7. Maximum *a posteriori* Probability: Not a criterion, but the best estimate of the probability that a signal plus noise was the process after the observation $x(t)$.

   $$P_e(SN) = \frac{l(x) \, P(SN)}{l(x) \, P(SN) + 1 - P(SN)}$$

To summarize the theory of the fundamental detection problem: all decision devices can be evaluated by considering the resulting operating point or operating curve on the ROC; the optimum decision devices are based on likelihood ratio, and operate on the upper boundary of the operating region on the ROC; specific operating points on the optimum operating characteristic are determined by specific operating criterion.

7.1.3 Fundamental Detection Problem, Applications

The essence of the theory is that the designer is assumed to know so much about the environment and the signals that are to be detected that he can actually specify how frequently various waveforms will occur. This is all wrapped up in the assumption that the probability densities $f_N(x(t))$ and

l_{SN} $(x(t))$ are known. By assuming varying degrees of uncertainty, or randomness, in the specification of the signals, and assuming various types of noise, the collection of solutions for special cases has grown over the past several years. If one desires specific results, he should look through the issues of the *Transactions of the Professional Group on Information Theory* of the IRE, and the technical reports of laboratories working in this area. The general results are briefly summarized here.

The likelihood ratio receivers fall into two categories. Those in the first category look remarkably like conventional receivers. Most of the pulse radar problems fall into this category. The reason for this is twofold: the assumption of stationary gaussian noise which is roughly white over the frequencies of interest coupled with a fairly well specified signal means that the logarithm of the likelihood ratio is proportional to the cross-correlation of the return and the signal, or some average of this cross-correlation over r-f phase, time of return, etc.; for pulse or other short duration local-in-frequency signals, the cross-correlation can be mechanized by passive filtering, and the averaging accomplished by video filters, CRT screen persistence, or similar common methods. Receivers in the second category look like special purpose digital computers. These result when the signals or the noise is of a more exotic (realistic?) type than for the first category.

The efficiency of the likelihood ratio receivers drops as less is known about the signal, and parameters are assumed to be random variables. This loss in efficiency is most pronounced at low energy levels, and usually is negligible at very high energy levels. By studying a series of such special cases where less and less is assumed known about the signal, one can gain a theoretical insight as to the importance or unimportance of knowledge of the various signal parameters. Such a general insight may be the most useful application of the theory, especially when the actual physical problem is too rough for direct solution, and the simplifying assumptions necessary to make the problem solvable were so strong that the resulting "ideal" receiver would fail to operate in the physical environment.

When situations are being studied in which the signal can be specified to a high degree of certainty, and the interference can be specified accurately, the direct application of the statistical analysis may disclose new equipment techniques that can be used to construct highly efficient receivers. In general, however, such a mathematical analysis can yield only guidelines, and some degree of structuring and ordering to the complex physical problem of optimum design.

7.1.4 Other Detection Problems

Two other areas in detection theory must be mentioned. These are the multiple decision problem, and sequential detection.

In theory, the multiple decision problem, deciding which of a specified number of mutually exclusive and exhaustive causes was the true cause of an observation, is not much more difficult than the special case of two causes discussed as the fundamental problem. When there were two causes and two alternative responses, there were four combinations of these. However, only two "relevant" probabilities were considered, $P_{SN}(A)$ for detection and $P_N(A)$ for false alarm, since the other two can be determined from these. The relation between these two formed an ROC plot, and all criteria for "best" fell on one optimum curve. When one has n possible causes and n possible responses, there are n^2 combinations and these reduce to only n^2-n relevant probabilities. Either some special definition of "best" is assumed, such as least-squared-error, or all errors are equally bad, in order to simplify the processing of the observation required of the receiver, or the resulting form of the receiver is n or $n-1$ parallel processors feeding into a comparitor device to arrive at the final decision. Thus, the essential simplicity of the fundamental problem, one input, one output, is immediately lost in the step from two causes to three. This increases the tendency of the receivers to be specified as special purpose digital computers.

Sequential analysis (References 21 and 22) can be applied to both the two-cause and the multiple-cause problems. The idea is that for the two-cause problem, three alternative responses can be made, the new response being a non-terminal decision to observe further. This is ideally suited for situations in which the signal continues in time, and for which a saving in time necessary to reach a certain quality of decision is good. In such situations the actual duration of observation varies, but often averages only at half the value necessary for a preselected observation time. Occasional very long observations also occur.

These extensions of decision theory have been touched here to remind the reader of, or alert him to, their existence. This is not meant to slight in any way the extensive work done in this area. Almost all of this, as well as the work on the fundamental problem, is available in the unclassified literature. The intent here has been to paint the field of detection with a very board brush to give the reader some feel for the field.

7.2 Methods of Signal Analysis

The main purpose of this section is to discuss briefly those methods of mathematical signal analysis that have been used in the past in various fields, including electromagnetic (EM) reconnaissance. The methods discussed can be instrumented, although in most cases a human observer is required for interpreting the results of measurement. By using different physical apparatus the observer can obtain a different and sometimes con-

densed representation of an analyzed signal. Certain significant signal properties may appear in practice more clearly in one representation than in another. It is not the purpose of this section to discuss the possibility of integrating these methods into the practical realization of an unattended signal analyzer system.

7.2.1 Decomposition of Signal Class

The signal at the input to the receiver will have the generalized functional form,

$$e(t) = A(t) \cos \theta(t) \tag{7-6}$$

where $A(t) =$ amplitude modulation (AM) function
  $\theta(t) =$ angle modulation function
In Eq (7-6) when $A(t) = A_o =$ constant and

$$\theta(t) = \theta_f(t) = 2\pi \int_{t_o}^{t} [1 + F_f(t)] f_o \, dt$$

$$= 2\pi \left[f_o t + \frac{\theta_0}{2\pi} + f_o \int_{t_o}^{t} F_f(t) \, dt \right] \tag{7-7}$$

we have pure frequency modulation (FM). When

$$\theta(t) = \theta_p(t) = 2\pi \left[f_o t + \frac{\theta_0}{2\pi} + CF_p(t) \right] \tag{7-8}$$

we have phase modulation (PM). In Eq (7-6), (7-7) and (7-8),

  $F_f(t) =$ FM intelligence
  $F_p(t) =$ PM intelligence
  $f_o =$ carrier frequency
  $\dfrac{\theta_0}{2\pi} =$ constant phase angle
  $C =$ proportionality constant
  $A_0 =$ amplitude of carrier
  $A(t) =$ AM function

When,

$$\theta(t) = 2\pi \left[f_o t + \frac{\theta_0}{2\pi} \right] \tag{7-9}$$

we have, pure AM. A device having a linear frequency response will give
an output

$$E_f(t) = f_o \int_{t_o}^{t} F(t) \, dt \qquad (7\text{-}10)$$

when FM is fed into it and, when PM is fed into it we have,

$$E_p(t) = MF(t) \qquad (7\text{-}11)$$

Hence, this device will respond to both FM and PM. However, in order to
recover the exact modulation in FM the output in Eq (7-10) must be differ-
entiated. This requires a filter whose spectrum increases linearly with fre-
quency. For the purpose of EM reconnaissance we can group FM and PM
into one classification and call it angle modulation. The most general type
of received signal will have the functional form,

$$e(t) = A(t) \cos \left\{ 2\pi \left[f_o t + \frac{\theta_0}{2\pi} + f(t) \right] \right\} \qquad (7\text{-}12)$$

where for FM

$$f(t) = f_o \int_{t_o}^{t} F(t) \, dt \qquad (7\text{-}13)$$

and for PM

$$= MF(t)$$

Eq (7-12) is the generalized mathematical representation of a transmitted
signal. If we neglect disturbances in the communications link between trans-
mitter and intercept receiver, then Eq (7-12) is the generalized mathematical
form of the signal at the antenna input. In Eq (7-12) it is seen that the
received signal can be represented symbolically by a four-dimensional point,

$$e(t) = \left\{ A(t), f_o, \theta_0, f(t) \right\} \qquad (7\text{-}14)$$

For completeness, we should also include the space time coordinates of the
source relative to the receiver as,

$$\Gamma = \left\{ \begin{array}{cccc} R, & \theta, & \phi, & T \end{array} \right\} \tag{7-15}$$

(T is the time at which signal is received.) All the information that can be obtained in EM reconnaissance from a single observation can be represented by an eight-dimensional parameter point. That is, there are eight domains from which information can be extracted. In EM reconnaissance, the arbitrary carrier phase angle θ_s is generally of no interest. Hence, we have a seven-dimensional parameter point.

$$e_{\Gamma}(t) = \left\{ \begin{array}{cccc} A(t), & f_o, & f(t), & \Gamma \end{array} \right\} \tag{7-16}$$

which represents all the information at the transmitter which ideally exists at the receiver. Physical and practical constraints at the receiver permit only a portion of the information to be extracted. However, we emphasize again that the detailed time structure of each intercept is of no interest in EM reconnaissance unless *a priori* information exists which justifies such an analysis.

The function of the EM reconnaissance observation is to operate on $e_{\Gamma}(t)$ in such a way so as to extract the required information in an optimum manner relative to a set of physical and practical constraints imposed on the observation and relative to the *a priori* information that is available to the observer.

For the purpose of our discussion, Γ is assumed constant. We also assume that the receiver is tuned to some fixed carrier frequency f_o and that it has the ability to reproduce the modulation functions $A(t)$ and $f(t)$. Thus, for the purpose of this discussion, we postulate that f_o and Γ are fixed points in our space and that observations must be made first to determine if there is signal present at the fixed coordinate, (f_o, Γ) and if so what the properties of the intercept are.

The methods of detecting the presence or absence of a carrier are discussed in the earlier portion of this chapter. This leaves the problem of measuring the significant properties of the modulation signals. However, a signal analyzer system is not entirely independent of its receiver nor is the receiver independent of the signal analyzer. For example, the receiver noise and bandwidth put physical limitations on the statistical uncertainties with which measurements can be made. Furthermore, if the receiver has sufficient frequency resolution so that the probability of more than one signal in a band is very small then there is no need for multiple signal separation on a video basis by the analyzer. Also, unity probability of intercept by the receiver does not mean that there is unity probability of analysis by the analyzer.

We assume that the presence of a carrier has been detected at a given frequency f_o and at a given coordinate of space time, (Γ, T). The next most logical operation is to study the properties of the envelope and of the angle modulation. In particular, we would like to know if there exists envelope modulation, or FM or perhaps both, i.e. we wish to know if,

$$A(t) = A_o = \text{a constant}$$
$$\qquad = a(t) = \text{time varying function} \qquad (7\text{-}17)$$

for the case of AM and for the case of angle modulation

$$f(t) = f_o = \text{constant}$$
$$\qquad = F(t) = \text{time varying function} \qquad (7\text{-}18)$$

In order to extract this information, we must be able to define one or more physical operators which yield a set of numbers from which the truth of the statements can be deduced. That is we define two operators R_A on $A(t)$ and R_f on $f(t)$ which yield the following results:

$$R_A[A(t)] = \lambda_A \qquad (7\text{-}19)$$

$$R_f[f(t)] = \lambda_f \qquad (7\text{-}20)$$

Then, the results of measuring the presence or absence of AM or FM modulation can be expressed as:

$$G[\lambda_A, I(\lambda_A)] = 1 \qquad [\lambda_A \in I(\lambda_A)]$$
$$\qquad\qquad = 0 \qquad (\lambda_A \notin I(\lambda_A)) \qquad (7\text{-}21)$$

and

$$G[\lambda_f, I(\lambda_f)] = 1 \qquad [(\lambda_f \in I(\lambda_f)]$$
$$\qquad\qquad = 0 \qquad [\lambda_f \notin I(\lambda_f)] \qquad (7\text{-}22)$$

Eq (7-21) states that AM modulation is present if the results of the operator R_A is an element of the interval $I(\lambda_A)$ and AM is not present if this is not the case. A similar interpretation is given to Eq (7-22). We assume that these measurements are simultaneous. All the possible results of measurement are defined by a two-digit binary number. For example, the binary number 10 implies the presence of AM and the absence of FM while 01 implies the presence of FM and absence of AM. Similarly the binary number 11 implies the presence of both AM and FM whil 00 implies the absence of both.

If it is established that envelope or angle modulation (or both) is present then it is necessary to decompose the class of envelope modulated and angle modulated signals further by defining more operators. However, from this point on the operations on the intelligence signals $A(t)$ or $F(t)$ can be the same. This should be clear, since both represent the intelligence that was transmitted, i.e. the same intelligence can be transmitted by varying either the frequency or envelope of the carrier. In EM reconnaissance we establish which parameter of the carrier is modulated by the operators R_A, R_f. However, beyond this point we may operate on the intelligence received. The operations should be independent of the type of r-f carrier modulation used.

Let us assume that it has been established that envelope modulation $A(t)$ is present. The next appropriate question is whether the modulation is pulsed or c-w. We can express this measurement operationally by,

$$R_1 A (t) = \lambda_1 \qquad (7\text{-}25)$$

The result of measurement as,

$$
\begin{aligned}
X_g(\lambda_i) I(\lambda_1) &= 1 \qquad [\lambda_1 \in I(\lambda_1)] \\
&= 0 \qquad [\lambda_1 \notin I(\lambda_1)]
\end{aligned}
\qquad (7\text{-}24)
$$

Eq (7-24) states that if the result of an operation on the signal $A(t)$ yields λ_1 belonging to the interval $I(\lambda_1)$ then c-w is present and if the result is not a member of $I(\lambda_1)$, then c-w is not present or a pulse signal is implied. The operator R_1 can be described as a measurement of duty ratio. If it is established that the carrier is pulsed, then it appears reasonable to decompose the class of pulse signals into subclasses such as PCM, PAM, PWM, PPM etc. All of these properties must be defined precisely by a physical measurement whose operational representation has the same functional form as Eq (7-23) and (7-24).

Each subclass of signals can be decomposed, at least in principle, until there remains only one member in each class. Once this has been accomplished the structural information of all the possible intercepts is known. The very large number of signals encountered in EM reconnaissance, make it impractical to store all the detailed characteristics of the signals. In most cases even if this information were available an observer could not hope to assign a distinct meaningful event or cause to each signal received. Relatively broad categories are a significant starting point. The problem of defining other operators with respect to a subclass of signals is simpler than defining them on the entire class of signals. Although the operations are discussed sequentially, they can be applied simultaneously. The important problem is to establish

a set of measurements or observables that are meaningful with respect to the hypotheses that are to be tested and with respect to the *a priori* information available. In the next section we examine some physical operations which define the signal classes discussed.

7.2.2 Low-Order Statistical Signal Analysis

In terms of the general ideas presented, we can define a "low-order" automatic statistical analyzer system for order-of-battle EM reconnaissance. The
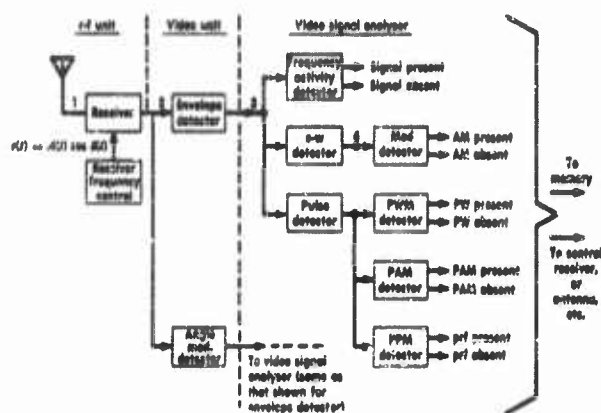
FIGURE 7-2 Functional Block Diagram of Low-Order Automatic Signal Analyzer

reader may visualize an actual equipment such as that shown in Figure 7-2. For low-order analysis it is essential to define few but significant subsets into which the class of all signals can be decomposed. It is important to choose a physical observable which is significant for the purpose of realizing this type of classification. There is no unique way, in the mathematical sense, of making this choice. The observer based on *a priori* information available to him must decide. The physical process of making the measurement is, in principle, capable of a mathematical definition.

Let us illustrate this important point with an example. Consider the problem of detecting a missile in space by means of a pulsed radar. The fact that a measurement at the receiver indicating a signal return implies a missile, is a definition established *a priori* by the observer based on information available to him. However, the problem of recognizing the presence or absence of a pulse signal in the noise involves a physical measurement which can be formulated mathematically. An optimum and automatic method of performing the measurement and of deciding whether signal is present in the noise or not can be realized.

Similarly, a measurement indicating that a received intercept is pulsed of constant amplitude, width and prf, is a physical measurement and hence capable of mathematical definition. However, the conclusion that the source of the signal is a radar is not capable of mathematical definition. This comes about as a result of a correspondence established by the observer between this type of signal and a given radar system. Another observer having different *a priori* knowledge can conceivably assign to the origin of this signal another source based on the same data. We therefore conclude that in EM reconnaissance as well as in all physical measurements there are essentially two sets of operations, one of which is chosen so as to relate a set of properties to certain events or situations while the other deals with the best way of processing the measured data, in order to make correct decisions.

We now proceed to define a physical observable which contains information on modulation properties. Let $x = (x_i)$ $i = 1, 2, \ldots, N$ be a sequence of positive samples that are measured on some property of a signal. For example, the sequence (x_i) could represent a measurement of the sample values of a c-w-AM carrier, or the pulsewidth intervals of a pulsed carrier, or the time between pulses of a pulsed carrier, etc. We define the new sequence of $(N - 1)$ samples,

$$y_i = \frac{x_i - x_{i+1}}{x_i + x_{i+1}} \qquad (7\text{-}25)$$

$$x_i \geqq 0 \text{ for all } i; \, i = 1, 2, \ldots, N.$$

The sequence (y_i) has the following desirable properties whether the (x_i) are statistically independent or not:

a) $y = F(x) = F(ax)$;

(y is independent of a scale factor on x)

$$E(y) = \lim_{N \to 0} \frac{1}{N} \sum_{i=1}^{N} y_i = 0$$

(The average value of y is zero)

c) $-1 \leqq y \leqq 1$ for $0 \leqq x$

(The semi-infinite range of x has been compressed to the interval $(-1,1)$.)

d) For the case where x is normal and the (x_i) are pairwise independent, the probability density function of

$$r(\tau) = \frac{\sqrt{2y}}{\sqrt{\dfrac{R(0)}{\mu^2} + \dfrac{R(\tau)y^2}{\mu^2} + \left[\dfrac{R(0)}{\mu^2} - \dfrac{R(\tau)}{\mu^2}\right]}} \qquad (7\text{-}26)$$

is normal with mean zero and unit standard deviation.

$R(0) = \sigma^2 =$ mean square value of variable x

$R(\tau) =$ autocorrelation function of $x(t)$

$\tau = t_1 - t_2 =$ sampling period.

$\mu =$ mean value of x

Therefore,

$$m = \frac{\sigma}{\mu} = \text{rms percent modulation.} \qquad (7\text{-}27)$$

The value at which $R(\tau)$ is evaluated depends on the sample spacing. When $R(\tau) = 0$, $(\tau > 0)$ the sampling is independent. Then,

$$r = \frac{y}{m}. \qquad (7\text{-}28)$$

Thus, the parameter of the distribution $P(y)$ is the rms percent modulation m and the statistic y has the desirable properties for studying the modulation characteristics. For the case where the statistics of x are not gaussian and pairwise independent, the measurement of y is still significant except that the analytical difficulties in calculation are insurmountable. When all samples have the same values the elements of $\{y_i\}$ are all zero, as required. This statistic yields some information about the order of the samples. Sign information in $\{y_i\}$ could also be useful for recognizing monotonic type modulations such as a VT fuse approaching the intercept receiver, fading characteristics, etc.

This statistic can be incorporated into a low-order analyzer system to indicate the presence or absence of a given type of modulation. For example, if $\{x_i\}$ represents the pulsewidth intervals of a pulse signal, then $\{y_i\}$ represents the pulsewidth to pulsewidth fluctuations. A similar measurement on amplitude will indicate if there is amplitude modulation present. If the $\{y_i\}$ are computed on pulsewidth, pulse amplitude and pulse repetition period and if each $\{y_i\}$ is then quantized to two levels then it is possible to detect eight combinations of modulations. The more levels into which the $\{y_i\}$ are quantized the more complex the system.

7.2.3 Signal Analysis By Statistical Methods

The signal intercepts at the output of a reconnaissance receiver can best be described by statistical methods. Receiver noise (plus other non-removable disturbance) even when combined with a deterministic signal such as a sine-wave of known amplitude, frequency and phase, yield an output function which can only be described statistically. Another important fact is that information generating processes are inherently statistical. For example, the basic sounds in speech have certain probabilities of occuring with respect to other sounds. In principle, all of the probabilities can be obtained by performing a sufficiently large number of experiments.

When a statistical process is at work, all of the information about the process is contained in an infinite number of probability distributions of all orders. If $X(t)$ is a typical member of an ensemble of random variables then the mth order distribution function, describing the process is given by (Reference 2):

$$F(X_1, t_1; \ldots; \ldots X_m, t_m) = P_r[X(t_1) \leqq X_1; \ldots; \ldots; X(t_m) \leqq X_m]$$

$$(7\text{-}29)$$

That is at time $t = t_1$, we sample all of the members of the ensemble and choose those which take on a value less than X_1. From these remaining members, we choose only those which take on a value less than X_2 at $t = t_2$. From the remaining members we choose only those that take on the value less that X_3 at $t = t_3$, etc. The relative number that remains gives the probability of the joint event expressed symbolically in Eq (7-29).

From this brief description it is clear that a measurement of any distribution of order greater than one becomes extremely complicated since the number of combinations of distinct measurements become very large. Furthermore, except for the gaussian process, where all orders of distributions can be obtained analytically, it is extremely difficult to obtain usable mathematical expressions for the high-order distributions. These serious limitations (both experimental and analytical) force us to simpler methods of describing statistical processes. The methods are incomplete of course, but are sufficient for most practical purposes encountered in communications.

The two most important descriptions of a statistical process are the first order distributions in conjunction with the power density spectrum or its equivalent the autocorrelation function. The first order distribution gives only the distribution of sample values and contains such important physical observables as the mean value (or d-c value) of the function and the mean-square value (or the power) contained in a signal. The power density gives the frequency distribution of the power in the signal. The autocorrelation function is the time equivalent of the power density spectrum. These func-

tions represent a specific statistic of the second order distribution and in general do not completely specify the second order distribution. From a practical signal analysis point of view it is convenient to look at the same thing in two different ways and hence measurements of the power spectrum and autocorrelation function might be desirable. Certain signal properties might also appear more readily on one of the representations.

There are two basic assumptions about the class of signals which must be made clear in order to be able to apply the statistical theories to physical problems. Fortunately, these assumptions are satisfied reasonably well for the class of signals which we encounter. First, we assume that the statistical processes we deal with are stationary. For example, the probability distributions of all orders remain invariant under a time translation of the random process. The second postulate is the "ergodic hypothesis". That is, we assume that statistical averages over the members of the ensemble are equivalent to time averages over a typical member of the ensemble. Without this assumption we could not perform statistical measurements in EM reconnaissance since we encounter only typical members of various ensembles. As a consequence of the ergodic hypothesis we can also subdivide a given signal into sufficiently long subsections each of which has approximately the same statistics. (We can therefore generate or approximate an ensemble from a long observation of a typical member). It follows, that if observations are taken sufficiently far apart in time the sample values are almost statistically independent.

Let $\triangle G_{X_i}$ be an operator on a typical member of a stationary ergodic ensemble of functions such that,

$$\triangle G_{X_i}[f(t)] = \triangle G_{X_i}(t) = 1 \text{ for all } t \text{ for which } X_{i-1} < f(t) \leqq X_i$$
$$= 0 \text{ for all other } t \qquad (7\text{-}30)$$

Then,

$$\triangle \lambda(X_i) = \lim_{T \to 0} \triangle \lambda_i(T) = \lim_{T \to 0} \int_0^T \triangle G_{X_i}(t) \, dt \qquad (7\text{-}31)$$

where $\triangle \lambda(X_i)$ is the theoretical probability that a sample X of $f(t)$ will be in the interval (X_{i-1}, X_i). If we quantize the entire range of $f(t)$ into N intervals, then we obtain, the discrete probability density,

$$\left\{ \triangle \lambda(X_i) \right\} i = 1, 2, \ldots, N$$

such that,

$$\sum_{i=1}^{N} \triangle\lambda(X_i) = 1 \qquad (7\text{-}32)$$

By the proper limit process, we can go from the discrete density to the continuous probability density if $f(t)$ has the appropriate mathematical properties.

The empirical distribution which is the one measured experimentally is simply the set of numbers, $\{\triangle\lambda_i(T)\}i = 1, 2, \ldots, N$ such that,

$$\sum_{i=1}^{N} \triangle\lambda_i(T) = 1 \qquad (7\text{-}33)$$

At a given set of defined quantization levels, the empirical distribution will converge to the theoretical distribution almost always. For the case where the function of $f(t)$ is sampled at discrete time points we obtain for the probability the expression equivalent to Eq (7-31) or,

$$\triangle\lambda(X_i) = \lim_{K \to \infty} \triangle\lambda_i(K) = \lim_{K \to \infty} \frac{1}{K} \sum_{j=1}^{K} \triangle G_{X_i}(t_j) \qquad (7\text{-}34)$$

Therefore, the measurement of the first order distribution requires quantization of the range of the variable, sampling (discrete or continuous) and counting (summation or integration). These measurements can be made by an on-the-spot analyzer and generally yield a significant amount of information, in particular when the shape of the distribution is preserved. The order of the sample values is not preserved. However, this measurement can be made on pulsewidth, pulse amplitude, prf and for that matter at the output of any linear or nonlinear operation. Figure 7-3 is a block diagram of a first order probability density function analyzer.

The autocorrelation function of a stationary ergodic process is given by Reference 2.

$$\phi_{ss}(\tau) = \overline{X(t)X(t+\tau)} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} X(t)X(t+\tau)\,dt \qquad (7\text{ }35)$$

where $\overline{X(t)X(t+\tau)}$ is the ensemble average. This function which is a second order statistic (for example, the mean value of the second order
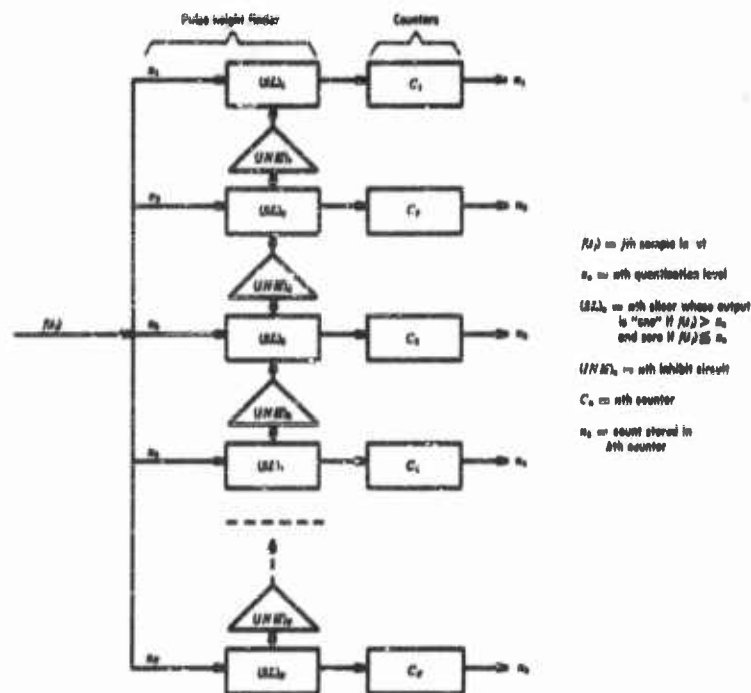
FIGURE 7-3   First Order Probability Density Analyzer

distribution) can be measured empirically although with more difficulty than the first order distribution. For a gaussian process, the first order distribution and the autocorrelation function determine all orders of probability distributions.

Very often the function $X(t)$ is sampled and the autocorrelation function of the sequence of samples is measured. This is given by,

$$\phi_{ss}(k) = \lim_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} X(n)X(n+k) \tag{7-36}$$

Figure 7-4 is a block diagram of a typical correlator. The autocorrelation function and the power density spectrum are Fourier transforms of each other. Hence, a knowledge of one implies the other. Thus, we have,

$$G_{ss}(\omega) = \frac{2}{\pi} \int_{0}^{+\infty} \phi_{ss}(\tau) \cos \omega \tau \, d\tau \tag{7-37}$$
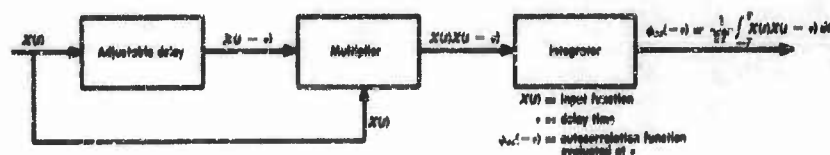
FIGURE 7-4   A Serial Analog Autocorrelator

and

$$\phi_{ss}(\tau) = \int_0^\infty G_{ss}(\omega) \cos \omega\tau \, d\tau \qquad (7\text{-}38)$$

The empirical measurement techniques of the power density spectrum are well known. We require a filter bandwidth which is much narrower than the signal bandwidth and a power measuring and averaging device at the output. The measuring filter must be capable of frequency translation over the bandwidth of interest. A search receiver is a good example of a spectrum analyzer. Of course, a bank of parallel filters distributed over the band of interest can also be used.

The operators defined by correlation and spectrum analysis also make sense with respect to deterministic signals. We can use these operations to gain some insight as to the time structure of an input without observing the entire function. It should be emphasized that autocorrelation and power-spectrum analysis destroy phase information. For more detailed analysis we can compare functions by cross-correlating them, for example, compute

$$\phi_{sy}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} X(t) Y(t + \tau) \, dt \qquad (7\text{-}39)$$

If $X(t)$ and $Y(t)$ are equal then this function will reduce to the auto-correlation function. In fact $\phi_{sy}(\tau)$ can be used to obtain a figure of "match" between $X(t)$ and $Y(t)$. The cross-power density spectrum is the Fourier transform of $\phi_{sy}(s)$. Both of these functions preserve phase information.

7.2.4 Summary and Conclusions

In this section we have formulated mathematically the observation or measurement problem applicable to EM reconnaissance. This formulation indicates that a major problem in EM reconnaissance is defining a set of significant measurable properties on the class of all signals. These measurements specify an important portion of an EM reconnaissance system and depend critically on *a priori* information.

We have discussed briefly a variety of physical measurements and apparatus which have proved useful in analyzing signals. The measurements are meaningful on both stochastic and deterministic signals. The physical constraints inherent in these measurements are more or less the same, for example, finite observation time, finite number of parameters, finite power, etc. Based on the particular application some of these measurements or operations are preferable to others. We indicated that for low-order analysis simple operators which yield the presence or absence of certain properties are useful. In particular, an analyzer indicating the presence or absence of certain types of modulation yields useful information.

REFERENCES

1. North, D. O., "An Analysis of the Factors which Determine Signal-Noise Discrimination in Pulsed Carrier Systems," RCA Laboratory Rpt PTR-6C, 1943.

2. Lawson, J. L., and G. E. Uhlenbeck, *"Threshold Signals,"* McGraw-Hill Book Company, Inc., New York, 1950.

3. Marcum, J. I., "Statistical Theory of Target Detection by Pulsed Radar: Mathematical Appendix," Rand Corporation Report R-113, July 1, 1948.

4. Slattery, T. G., The Detection of a Sine Wave in Noise by the Use of a Non-Linear Filter, *Proc., IRE*, Vol. 40, p. 1232, October, 1952.

5. Hanse, H., "The Optimization and Analysis of Systems for the Detection of Pulsed Signals in Random Noise," Doctoral Dissertation (MIT), January, 1951.

6. Schwartz, M., "A Statistical Approach to the Automatic Search Problem," Doctoral Dissertation (Harvard)　ne, 1951.

7. Middleton, D., "Statistical Criteria for the Detection of Pulsed Carriers in Noise," *J. Appl. Phys.*, Vol. 24, p. 371, April, 1953.

8. Middleton, D., W. W. Peterson and T. G. Birdsall, Discussion of 'Statistical Criteria for the Detection of Pulsed Carriers in Noise, I, II', *J. Appl. Phys.*, Vol. 25, pp. 128-130, January, 1954.

9. Woodward, P. M., and I. L. Davies, A Theory of Radar Information, *Phil. Mag.*, Vol. 41, p. 1001, 1950.

10. Woodward, P. M., Information Theory and the Design of Radar Receivers, *Proc. IRE*, Vol. 39, p. 1521, 1951.

11. Woodward, P. M., and I. L. Davies, Information Theory and Inverse Probability in Telecommunication, *Proc. IEE*, (London), Vol. 99, Part III, p. 37, March 1952.

12. Davies, I. L., On Determining the Presence of Signals in Noise, *Proc. IEE*, (London), Vol. 99, Part III, pp. 45-51, March 1952.

13. Grenander, U., Stochastic Processes and Statistical Inference, *Arkiv Mathematik*, Vol. 2, p. 195, 1950.

14. Reich, E., and P. Swerling, The Detection of a Sine Wave in Gaussian Noise, *J. Appl. Phys.*, Vol. 24, p. 289, March, 1953.

15. Davis, R. C., On the Detection of Sure Signals in Noise, *J. Appl. Phys.*, Vol. 25, pp. 76-82, January, 1954.

16. Davis, R. C., Detectability of Random Signals in the Presence of Noise, *Trans. IRE*, PGIT-3, p. 52, March, 1954.

17. Peterson, W. W., and Birdsall, T. G., "The Theory of Signal Detectability," Technical Report No. 13, Electronic Defense Group, Department of Electrical Engineering, The University of Michigan. This was essentially the same as the paper: Peterson, W. W., T. G. Birdsall, and W. Fox, The Theory of Signal Detectability, *Trans. IRE*, PGIT-4, September 1954.

18. Middleton, D., "The Statistical Theory of Detection, I: Optimum Detection of Signals in Noise," MIT Lincoln Laboratory, Technical Report No. 35, November 2, 1953.

19. Middleton, D., Statistical Theory of Signal Detection, *Trans IRE*, PGIT-3, p. 26, March, 1954.

20. Urkowitz, H., Filters for Detection of Small Radar Signals in Clutter, *J. Appl. Phys.*, Vol. 24, p. 1024, 1953.

21. Wald, A., "Sequential Analysis," John Wiley & Sons, Inc., New York, 1947.

22. Blackwell, D., and M. A. Girshick, "Theory of Games and Statistical Decisions," John Wiley & Sons, Inc., New York, 1954.

8

Psychophysics in Electronic Warfare

W. TANNER, JR.

8.1 Introduction

Anyone engaged in a countermeasures program will recognize at once the complexity of the problem before him. He is faced with equipment which he must evaluate in terms of both theoretical and experimental considerations. Conditions vary with the environment, and evaluative techniques which are satisfactory in one particular environment may be next to useless in another. Besides, there is the added complication of the human factor. The exact performance of a human being cannot be predicted with absolute certainty, and one must take into consideration the unreliability of the human component in any relatively flexible system. These problems are not easily solved. Indeed, in the past there has frequently been disagreement over the question of how the solutions should even be approached. Should we approach the matter from the point of view of theory, carefully stating the problem, controlling the environment, and collecting laboratory data for the evaluation of countermeasures systems? Or should we proceed from a practical point of view, considering as well the irrelevant variables found in field tests for the evaluation of equipment under specific conditions?

The present chapter adopts the former approach for a variety of reasons. The distinction between theory and practice is sometimes more apparent than real for theoretical developments frequently have highly significant practical aspects. Such seems to be the case in the countermeasures prob-

lem. Evidence already gathered seems to point to the conclusion that only by approaching the matter from the theoretical point of view can the relevant relationships between the whole system and its parts be adequately studied. We wish to get to the fundamental nature of the problem, uncomplicated by irrelevant variables, so that we may both understand the functioning of a system and develop means for its evaluation. Such an approach requires that the subject be viewed in the light of recent communication theory.

Admittedly, such theory is extremely complicated and by no means easy to understand. It cannot necessarily be assumed to be in the general background of even very well-educated men. For this reason, this chapter assumes no knowledge of communication theory on the part of the reader. For those familiar with this area, the discussions which follow may seem overly simplified. The matter is handled in this way, however, so that any intelligent reader will be able to follow the discussion once he has become acquainted with the fundamentals of the theory. The chapter attempts to outline and explain the concepts from communication theory relevant to the present discussion in the hope that by this means the reader will come to see the significance of these new concepts for the evaluation of countermeasures programs.

A strong psychophysical bias will be noted at once. This orientation results from the primary field of interest of the author. The psychophysical viewpoint, however, is not the only, nor perhaps the most important one made in the subsequent discussions. In the countermeasures problem, we are faced with the task of evaluating both men and equipment. Fundamentally, the task is the same whether one or the other is evaluated in terms of the whole. As will be shown below, it sometimes makes no difference whether a component under discussion is human or mechanical. It is the function of that component in relation to the system as a whole which is of primary significance. The reader, therefore, should keep in mind that we are dealing with systems. The psychophysical matters, though important in themselves, are also useful in illustrating problems which have more general ramifications than psychophysics alone.

8.1.1 The Basic Problem

In any communication or detection system, two types of general problems are likely to appear. The first, and most easily recognized and understood, is the kind that relates to the particular components individually. Such problems may be isolated with more or less ease and referred to particular specialists for solutions. Thus, the problems caused by the individual human component may be referred to the psychologist, those of the elec-

tronic components to the engineer. In many instances, however, a second kind of problem appears—that which concerns the relation of the individual part to the system as a whole. Thus, to use an example from psychophysics, the inclusion of a human component in a communication or detection system introduces not only the problem of human limitations but also the problem of those limitations as they are related to factors outside the human being himself, most particularly, the lack of reliability of the human component in a flexible system. If the purpose of the equipment can be made specific enough, automatic equipment is perhaps better but in most systems that cannot be designed for a completely specific purpose, a human being must be included.

The exact nature of functional structure of this human component depends, in part, upon the way in which it is incorporated into the system. Human beings are extremely complex creatures, and unlike electronic equipment, cannot, at least at present, be represented by a schematic diagram of filters, amplifiers, differentiators, or the like. To be sure, there are specific tasks wherein the tie-in to the system and the purpose are so precisely specified that one fundamental schematic diagram of the human component may be used. If the task is changed, however, or if the tie-in to the system is changed, many of the parameters of the diagram most nearly representing the human component must also be changed. In some cases, it may even be necessary to change the basic functional structure represented in the diagram.

The underlying cause of this condition is that the human being is a self-evaluating and self-adjusting system—a fact which leads to both disadvantages and advantages. On the debit side, a number of major problems occur when an outside observer tries to predict the performance of a human being, particularly if the outside observer has difficulty estimating the precise environment in which the system is operating. Suppose, for example, he requires that the human component handle more factors than he is capable of, or demands a storage capacity on the part of the operator beyond the limits of human memory. Situations like these would seriously affect the efficiency of the system or lead to its malfunction, simply because the human being could not operate according to specifications. Thus, the dangers incurred with the inclusion of a human component are easy to visualize.

Yet, although these difficulties of prediction are major ones and can seriously affect the performance of the system, they are not entirely disadvantageous. On the credit side, the incorporation of a human being into a system can lead to real advantages. An intelligent component can adjust if the original estimate of the environment is inaccurate. He can change

with the situation by adjusting himself or the apparatus (or both) to fit the unforeseen conditions. Indeed, he can sometimes perform better than a mechanical component. Suppose, for example, that a mechanical device is set to receive a signal at precisely 1000 cycles per second. If the signal comes through at 1100 cycles per second, the system fails completely. A human being, on the other hand, can react to the *unexpected environment*. If there is time, he may adjust to the new situation, manipulate his equipment, and receive the signal at the transmitted frequency. His behavior may, in effect, change the design of the system. He may thus perform better than specific automatic equipment because he adds a flexibility, a means of adjusting to the unexpected, that is beyond the capacity of most machines.

Because an intelligent component can make up for system deficiencies, his introduction can be of real importance to the engineer. The inclusion of a human being may permit certain design decisions to be delayed until the system actually goes into operation; at that time the human operator may make the necessary adjustments to the specific conditions he finds. In certain cases, therefore, the design need not be made too specific, for, in time, the human component may adjust to the actual—and perhaps unpredicted— environment. He may even make the system work better than it would if specific design features had been incorporated. In effect, the system can deliberately be made less specific so that the human component may make changes when more information is available upon which to base a design decision.

Perhaps an example will aid in clarifying this point. It is drawn from the theory of signal detectability, and a number of the concepts and terms may well appear new and unusual. Nonetheless, it illustrates the point well, and the explanation will introduce the reader to some concepts that will be useful throughout the chapter. According to this theory, the cutoff point for accepting a signal (perhaps the bias on a thyratron) is optimally established on a knowledge of a number of parameters which define the particular environment in which the equipment is operating—parameters which can vary considerably from one situation to another. Stated flatly, these parameters are, among others the *a priori* probability of the signal occurring, the signal energy, and the values and costs associated with possible alternate decisions.

The way in which these variables play their roles can perhaps be demonstrated by the following illustration, which is not technical, but which is quite graphic. Suppose that at the end of a bus line frequently used late at night by nurses coming off the late shift at a hospital, there is a dark and deserted stretch of street lined with bushes and weeds. For many years, girls have walked along this stretch completely unmolested. Should one of

them hear a noise in the bushes, it would be accepted as a signal that a small animal is scampering away (the *a priori* probability—the probability before the event—is high that that is all it is). The girl is not frightened for it seems that little is to be gained by panic (value is low), and she may be laughed at (the cost is high) for running at every noise. However, suppose further that a new factor is added to the situation. One night a girl is murdered. Once this event occurs, the values of the parameters change. It is highly likely that many reports will come in of girls running home in panic pursued by a molester, perhaps imaginary. What has happened? Because of the murder, the *a priori* probability, in the minds of the girls, of the signal meaning "murderer" becomes high. That is, to them any noise is likely to denote "murderer lurking near." This signal is heard more frequently (though the noise is perhaps of only very low intensity), and the values and costs have changed. If the signal is accepted, as it is likely to be (that is, "murderer is there"), the girl runs because the value of such action ("escape with my life") far offsets any possible cost. A change has occurred in the system because of the new information (the murder) which has been introduced.

To return to the engineering example, the problem faced by the engineer in designing a system is now obvious. The parameters vary from situation to situation, just as they did in the nurse example. Therefore, in designing a fixed bias into a system, the engineer must assume an environment, or a set of environments, which leads him to elect that specific bias. But at the same time, obviously, there are likely to be many environments for which his choice is a very poor approximation. One new fact, as the murder of the nurse, fed into the system can change the parameters or their values so markedly that the system is totally unsuited for the environment in which it is asked to operate. And, of course, the engineer has no way of knowing what factors may crop up to bring about such a change. He may prefer, therefore, to provide for an adjustable bias in his system.

He would like to design a system which operates initially on the design engineer's best guess of the values of the relevant parameters in the environment. During its operation, the system studies the environment, making new estimates of these values based on the information it collects. It continually modifies its basis of operation to conform to the most recent estimates of the environment. Since at the present time, however, it is a difficult problem to design these properties into automatic devices, the engineer must assign some of the functions of the system to a human being. With present knowledge he can in no other way achieve the flexibility and versatility required. He assumes, therefore, that the human component can introduce into the system the necessary self-evaluation and self-adjustment

in a satisfactory manner. That assumption, however, is based on little evidence and much faith. Without question, the human component performs these functions. That he performs them satisfactorily is by no means so certain.

Until quite recently, this whole question of human performance had not been deemed sufficiently important to justify major consideration, mainly because it had not been necessary to demand the best possible use of human components in both communication and radar systems. For example, the relatively small amount of traffic in communication systems allowed plenty of bandwidth to be assigned to individual networks, and in radar systems, there has ordinarily been sufficient time for human beings to react to signals. Under such conditions, it was not at all necessary to demand optimum use of the human component. Recent developments, however, indicate that in the not-very-distant future, systems which employ human beings will have to make much greater use of human capabilities, mainly because of reduced time and increased traffic in radar and communication systems.

Consider, for example, the radar problem. The speed of modern missiles and counter-missiles is such that recognition and interception must occur in intervals of time beyond the fastest of human reactions. To use a commonly recognized illustration, the problem of the human operator is similar to that of two pilots in jet planes approaching each other while flying by visual observation alone. Speed has become so great and reaction time requirements so stringent that the planes may collide before the pilots are even aware of any danger. To be sure, these are extreme cases in which the unaided human being can do nothing. In other cases where more time is available, the human capabilities will have to be taken into careful consideration.

The problem exists as well in communication systems. Traffic has increased to such an extent that extravagant assignment of bandwidth to individual nets is no longer possible. The nets must operate on narrower bands, a condition which entails some serious problems, not all of which will be immediately apparent to the reader since they involve some concepts from information theory (they will be handled in more detail in Section 8.2). Nevertheless, the general nature of the problem may be mentioned here; more detailed analysis of the concepts will follow. It is hoped that as we proceed, the matter will become increasingly clear.

The first and most obvious effect of the narrower bands is the reduction of the information-carrying capacity of the nets. If the same rate of information flow is to be maintained, this effect entails the need for some efficient use of the assigned channels. A number of consequences come to mind. Suppose we assume a human being speaking through a channel. The code he uses may have to be more efficiently constructed in terms of the

capacity of the channel. (The word "code" is not used here in its most usual meaning. Any accepted means for transmitting information—language, flashes of light, dots and dashes, even tone of voice—may be classed as a code.) If the human voice employing language is transmitted over a narrow channel, the tone of voice or the limits of high and low inflections (and this is part of the code) that must be used if the message is to be received with minimum uncertainty may have to be specified. Thus, the codes must be carefully designed, for we cannot afford to depend upon communication systems for which our expectations are unrealistic.

More efficient utilization of the channel capacity will, of course, increase the susceptibility of the system to jamming. Jamming reduces the capacity of the channel. Since the bands are smaller, less noise power will be needed to jam the message (though naturally it will be harder to introduce). To counter this susceptibility, it will be necessary to make an intelligent use of redundancy. This means that those elements which do not add information should serve, in effect, as error-correcting elements, thus increasing the certainty with which the message is received. Notice, for example, the redundant letters in an Englis sentenc; we can reconize the words even if leters r omited. A wideband gives us a kind of wasteful redundancy—much more than we need. Since we will be forced to avoid this type, we will have to use that which remains to reduce the danger of equivocation—of uncertainty on the part of the receiver—which increases with the narrower bands. For example, detailed standard operating procedures may have to be specified to fit existing conditions so that redundancy is definitely built into the system.

In the future, therefore, it appears that more attention will have to be directed toward the problem of using the capabilities of the human component more efficiently. The knowledge basic to this consideration will come from psychophysical experiments, probably performed within the framework of models of developments so recent that their potential has not yet been generally realized—statistical decision theory and the theory of signal detectability. This area of research has only a brief history. It dates primarily from Shannon's paper on the mathematical theory of communication, published in 1947. This and other studies, as will be seen in subsequent

tions, have furnished the scheme of the basic communication system and certain fundamental concepts pertinent to the present problem. Although, at first glance the concepts seem difficult (the vocabulary, in particular, is not easy to master), they are of such importance that anyone working in a countermeasures program should be aware of what they imply. For despite its short history, the theory has already demonstrated a potential to furnish some of the required knowledge.

We shall begin, therefore, with a brief outline of the pertinent concepts

in information theory and analyze the communication, radar, and countermeasures problems in terms of these concepts. All three problems are included for a definite reason. Before the countermeasures problem can be fully understood, the reader must be thoroughly familiar with the natures of the target system themselves. Since the communication and radar analyses illustrate problems inherent in the target systems, each is treated at length before the analysis of the countermeasures system. The communication problem is basically a recognition one. It involves two people in a communication game in which the message ensemble can be known in advance. The radar problem differs in that it also involves detection, and the message ensemble (the targets) cannot be as precisely known in advance. Only when these problems are clarified can the countermeasures problem be properly analyzed.

Analysis of these problems will reveal a number of areas requiring research. Some of the problems are quite specific. Thus, in the communication problem, as will be seen below, this question arises: information already gathered indicates that from the decoding standpoint, sequential observation (observation over a flexible time interval until the observer achieves a level of confidence that he has received the message correctly) is more efficient than fixed time observation. Since this is true, it becomes important to know the extent to which a human operator can act as a sequential observer. Other areas for study include those in which even basic knowledge is lacking; for example, the manner in which a human operator on the receiving end of a communication channel stores and decodes messages (that is, reconstructs the transmitted message from the signal he has received).

Each of the three general problems (i.e., communications, radar, and countermeasures) is stated within the theoretical framework of communication theory and the theory of signal detectability. By establishing a single consistent framework it is possible to treat complex problems involving more than one of these systems. The framework leads to the establishment of those relevant measures upon which the evaluation of these complex systems and their components can be based. It is, further, the framework within which basic studies should be conducted for the purposes of increasing the scientific knowledge leading to future advances in the countermeasures program.

8.2 The Communications Problem

Before any effective countermeasures can be taken, the fundamental nature of the communications system must be clearly understood and the problems relevant to the general area of communications carefully analyzed.

The problems of concern here are inherent in the systems themselves, a fact which must be kept constantly in mind throughout the discussion. Effective countermeasures must therefore be based on a thorough understanding of the system. In this section, special problems of coding, decoding, and the use of redundancy will be analyzed with a view toward suggesting a number of areas in which further research and study are especially desirable. Since these terms, however, are used in a very special sense, a considerable amount of background material will have to be introduced.

The basic communications system and its important components are well illustrated in the block diagram upon which Shannon based his fundamental theorems, and the theorems themselves (although they are statements of averages which cannot be indiscriminately applied to realizable systems) are useful in the formation of the problems. A word of warning should perhaps be added here. The application of the theory depends upon large samples which permit the application of statistics. It is a theory of averages; although predictions can be made about a system *on the average*, the outcome of any individual event cannot be predicted. For example: the rate of radioactive decay is essentially a statistical phenomenon; the behavior of an individual particle cannot be predicted. In a similar fashion, in information theory, theorems are based upon a large number of messages; the theorems cannot furnish the basis of stating the fate of any one message in particular. This problem causes difficulty because individual messages are of concern in the application of information theory.

Shannon's diagram, modified only by the addition of a feedback channel, is reproduced in Figure 8-1. It is a general system not descriptive in detail
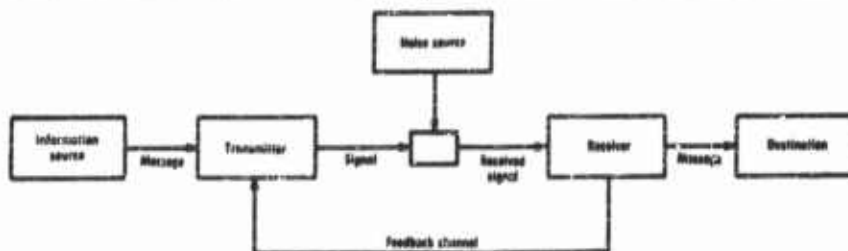
FIGURE 8-1   Basic Communications System

of any specific system. Nevertheless, its very generality is the basis for its usefulness, for it describes—abstractly—any communications system, from high-speed electronic devices to two friends engaged in conversation, from a radio or television broadcast to a pair of Indians sending smoke signals. In each case, the basic components are the same.

There is an *information source* which selects a desired message from a

act of possible ones, each with a probability that it will be transmitted when transmission occurs. The *transmitter* encodes it into a suitable code. That is, it changes the message into the *signal* which it sends over the comunications *channel*, any medium—a pair of wires, a band of radio frequencies, a beam of light—that can be used to transmit the signal to the *receiver*. The receiver accepts the *received signal*—the transmitted signal plus noise, any addition or change (distortion or error) not intended by the transmitter—and decodes it. That is, it reconstructs the message from the signal and passes it on to the *destination*, the person or thing for whom the message is intended.

Perhaps an example will illustrate the fundamental nature of the abstract system. Consider two friends engaged in a conversation. The speaker's brain is the information source; his vocal system, the transmitter. His words (the coded message) are transformed into varying sound pressures (the signal) and transmitted over the air (the channel). His friend's ear and auditory nerves are the receiver; his brain, the destination. Similar examples could be constructed from any situation in which communication is taking place. The general system describes them all equally well.

From the point of view of this section of the chapter, the most significant problems are those arising from the coding and decoding of the message. These problems are the same whether the coding involves language, binary digits, flashes of light, or current passing through a cable.

Around the block diagram Shannon constructs a coding theory, statistical in nature, which will be useful in the discussion of coding problems. To be sure, his theory must be used with care, for, as Shannon himself has noted, the hard core of the theory is a branch of mathematics, a strictly deductive system. The theorems he presents are statements of statistical averages expected of communications systems exhibiting stable statistical properties. This means dealing with infinite series of events, all of which have the same experience entering into them. The system is not altered by experience, nor does it change as it goes along, unless such change is built into the system by some general statistical rule. Let us also stress again that we are dealing with averages, which, like all averages, cannot be applied indiscriminately to particular situations. Shannon's fundamental theorem, for example, assumes transmission forever. Such a principle, based upon infinite time, cannot be applied to a realizable, finite system unless one is well aware of the dangers lurking in statistical statements.

Still another difficulty has to be mentioned. The vocabulary of information theory is somewhat hard to master, not because the words are difficult, but because they have common as well as specialized meanings. We have already noted the rather special sense in which the word *code* is used here

to mean something much less specific than in ordinary usage. Other words are even more difficult to grasp; information, for example, is *not meaning*, but rather a measure of one's freedom of choice when selecting a message. Confusion can result in one's mind because a word like information has these two meanings—one in general use, one in terms of information theory—both of which are needed in discussing communications. The reader must constantly keep in mind, therefore, how these words are being used.

Nevertheless, despite these limitations, Shannon's theorems are useful in helping to formulate the problems and will therefore be used throughout this section, especially in the analysis of the problems of coding. No attempt will be made to prove these theorems. The proofs require a considerable amount of mathematics and are clearly beyond the scope of this chapter. The interested reader may consult Shannon's work itself. For the purpose of this chapter, it will have to suffice to state and explain the theorems so that the reader may understand the concepts involved. Most pertinent to this discussion are the four concepts detailed below.

(1) *The information flow in a communication system is the same whether viewed from the standpoint of the transmitter or the receiver.* Again the reader must keep in mind, both here and throughout the entire discussion, that information, as used here, is *not meaning*, but rather a measure of one's freedom of choice when one selects a message. Perhaps the term can best be explained as follows: In information theory, the word *entropy* is used to express the degree of randomness in a situation. It is measured logarithmically and is expressed in terms of the probabilities involved. The more equal the probabilities, the greater the entropy. Information, then, is a measure of the reduction of entropy, or of uncertainty, in a communications system. What the theorem is saying in effect is that the channel can reduce uncertainty by so much, and it does not matter from which end you look at it.

Two examples may help to illustrate this concept. Let us assume a channel transmitting the digits 0 and 1, with each digit equally likely to be transmitted. From the transmitter's point of view, the rate of information flow is measured by the change in his ability to predict the *received digit* before he knows the symbol to be transmitted and after he knows it. From the receiver's end, the rate of information flow is the change in *his* ability to predict the symbol which is transmitted before and after the signal is received. In our assumed channel, the 0's and 1's may be mapped into two things each. Let us say that a 0 may be A or B, a 1 either C or D. 0 and 1 are equally likely to be transmitted, and the received digits are A, B, C, and D.

From the transmitter's point of view, before knowledge of the signal to be transmitted, the receiver will receive one of four digits, each with a probability of .25. With knowledge of the signal to be transmitted, the receiver will receive one of two digits, each with a probability of .5. The uncertainty has changed from two bits to one bit; it has been reduced by one bit. From the receiver's viewpoint, before knowledge of the signal, either a 0 or a 1 will be transmitted, each with a .5 probability. There is, therefore, one bit of uncertainty; it has been reduced by one bit. Now notice what has happened. The transmitter's uncertainty has been reduced from two bits to one bit, the receiver's from one bit to none. In each case uncertainty has been reduced by one bit. Since the information flow is measured by the reduction of uncertainty, it is clearly the same no matter which end of the system is considered.

Let us examine another example. Consider a system transmitting the binary digits 1 and 0. For the sake of a simple illustration, we will assume that it is a symmetrical system. The probability before selection of the digits is .5 that either will be selected. This probability is the same from the point of view of either transmitter or receiver. Suppose further that once a 0 is transmitted, the probability that a 0 will be received is .9 from the transmitter's point of view. After the perturbed signal (signal plus noise) is received, the probability that the 0 is correct is .9 from the receiver's point of view. For ease of calculation, we have assumed a symmetrical system. The same holds true for an asymmetrical system. The calculations, however, are so involved as to be out of place in this book. Actually, the point should be clear without them: the channel can reduce uncertainty by so much, and it matters not at all whether it is viewed from the transmitter's or the receiver's end.

(2) *There exists a code which leads to arbitrarily small error and still permits information flow at a rate nearly that of the capacity of the channel, whereas any code which attempts to transmit at a greater rate must lead to an equivocation* at least as great as the difference between that rate and the capacity of the channel.* This means that so long as the capacity of the channel is equal to or larger than the entropy of the source of the messages, we can transmit with as small a specification of error as we choose, so long as the transmission process continues over an infinitely long period of time. With time unlimited, error-checking devices can be built into the system. The longer the time allowed for this purpose, the more effective these checking devices can be.

Consider some of the means by which this can be accomplished. One method of error correction is simple repetition. The message, for example,

---

*Equivocation refers to the average uncertainty existing after the transmission.

may be repeated once—1389, 1389—giving a redundancy of 50 percent in the total message. Or the message might be repeated three times—2469, 2469, 2469 (redundancy is 67 percent). We might prefer to take the sum of the digits, 21, (redundancy is 33 percent), or perhaps only the last digit of the sum, 1, (redundancy is 20 percent) to show that the message is correct. The principle is one familiar to those who have had experience with legal language. In an act of sale of a piece of property, a lawyer may write that the seller agrees to "grant, bargain, sell, convey, assign, transfer, and set over" that property to the buyer. What he is doing in effect is building redundancy, an error-checking device, into his message to make sure that it is not misunderstood. It can easily be seen, therefore, that with infinitely long time for transmission, messages can be sent with arbitrarily small error, so long as the capacity of the channel is equal to or larger than the entropy of the source of the messages.

The above examples are merely intended to illustrate the role of redundancy. In the first example, that of repetition, no new information is transmitted after the first transmission. The second example again has this feature; the transmission of the sum is pure redundancy and serves only to detect errors. In the third, that of the legal language, each word carries some of the information included in every other word and some that is not. A single word could be coined to cover this set of words. None of these examples pretends to illustrate a use of redundancy which reduces error without reducing the rate of information flow.

If, on the other hand, one attempts to transmit at a rate greater than the capacity of the channel (the capacity in this case is smaller than the entropy of the source), an equivocation would be incurred equal to or greater than the difference between the capacity and the rate. Of main concern here is the relation between the capacity of the channel and the rate of transmission. Thus, if the channel capacity is varied (as when employing countermeasures) so that the channel capacity is less than the rate of transmission, the theorem states that there must be equivocation, or uncertainty, when the message is received.

The remaining concepts may be stated very briefly:

(3) *The rate of information flow from transmitter to receiver in the system illustrated in Figure 8-1 is restricted only by the capacity of the forward channel.*

(4) *The feedback channel can be employed to reduce the error rate of the information flow from transmitter to receiver, but not to increase the capacity of the forward channel.*

Other concepts, especially some from statistical decision theory, will also be employed, but since they are related primarily to problems in decoding, they will be discussed in Section 8.2.2.

8.2.1 Coding

From the discussion above it is obvious that the employment of redundancy is an important coding problem. A great deal of highly sophisticated mathematical effort is being extended in this direction. The fact that this is not treated in detail in this chapter should not minimize the importance of this work. It is extremely important, but requires more technical discussion than is possible in this chapter. Consequently, the discussion is devoted to voice-communication channels, or subject matter not sufficiently understood to permit sophisticated mathematical treatment.

A number of coding problems deserving further study have been suggested by research conducted at The University of Michigan. Two of these problems in particular involve the use of a human being on the transmitting side of a communications system. These are interrelated, for they involve not only the question of coding procedures in voice communication so that the greatest information rate per unit time may be achieved, but also raise the question of how these procedures should be determined for a given situation. It has already been shown that a human being can add flexibility to a communications system simply because he is a self-evaluating and self-adjusting component. To apply this concept to the coding problem, one may note that he has the ability to change the code through the use of varied reading rates, changes in volume, shifts in inflections, and the like. The problem, therefore, is how to put this flexibility to use. In other words, what can the human operator do to adjust his coding procedures to achieve the highest information rate per unit time? And more important, is it possible to devise a general rule which will tell the human operator which of the changes to use in a given situation?

From this point of view, it is important first to know the maximum information rate achievable in any given system. Figure 8-2 illustrates this problem. It is derived from calculations which are beyond the scope of this chapter. The reader is asked, therefore, to accept the graph as a true picture of the situation. The graph shows that, for a fixed signal-to-noise power ratio in a digital channel, the information rate per unit time varies as a function of the energy per symbol, or equivalently, as the duration of the symbol. It applies to a channel in which each symbol in a sequence is independent of every other symbol in the sequence. The two lines illustrate the relation in situations based upon two different assumptions. The first assumes that the signal is known exactly; that is, if the signal exists, its waveform is known exactly, even to its position in time. The curve shows that the reduction of
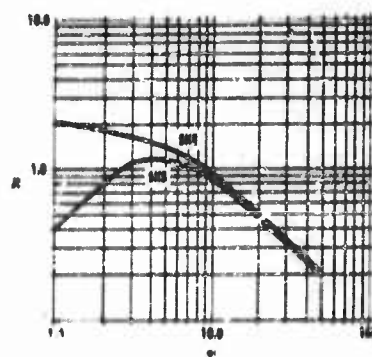
FIGURE .   Rate of Information per Second as a . . . of Signal Energy when Signal-Noise Power Ratio Is Constant

error in such a channel can be accomplished by increasing the energy of the symbol, in this situation, only at a cost or a loss of information rate.

The other assumes that the signal is known only statistically; that is, that the waveform is known, but perhaps not its position on a fixed time scale, or perchance, the amplitude. In such a situation, the curve achieves a maximum point in time, a fact which suggests the need for further study in coding procedures. Is it possible, in a voice-communication channel, to achieve something like this maximum? Is it possible to plan the redundancy in a way which permits the reduction of error without paying an excessive price in reduction of information flow? According to Shannon's fundamental theorem, there exists a code which can reduce error to an arbitrarily small amount without the loss of information rate illustrated in the first curve. By this theorem, also, there exists a code which can lead to an approximation of the information rate indicated by the maximum of the second curve in Figure 8-2, with arbitrarily small error. Although it would be extremely optimistic to attempt to achieve these codes in voice-communication channels, their existence indicates directions in which progress might be sought.

This optimum code, however, requires infinitely long delays to achieve the arbitrarily small error rate, for the theorem assumes transmission over an infinite period of time. In situations in which a fixed amount of information is to be transmitted, or in which the messages are independent, the theorem may not apply. In such situations, correction must come within the message itself or not at all, for one frequently cannot afford the time needed to assure minimum error through correction devices built into the code. If the message contains information urgently needed at the receiving end, we cannot waste time purely for the sake of an arbitrarily specified reduction of error. Nevertheless, there may well be optimum codes achieving a minimum error rate in accordance with the restrictions placed on the communications task. In other words, of a number of possible codes, we would like to find the one which, though falling short of the ideal, still has a minimum error rate in terms of the finite task that the communications system is required to perform.

For situations in which the signal is known only statistically, it becomes

highly important to conduct studies, either theoretical or experimental, to establish maxima, or approximations of maxima, for curves such as that in Figure 8-2. When a human component is built into the system, such studies would include research in voice communication. One would like to know, for example, the effect of reading rate in the establishment of such a maximum point. Other variations, such as change in volume, tone, and pitch in the human voice should be considered as well.

An example may clarify the problem. Suppose that a babble of voices is used as an interference signal in a communications system. When this sound is evaluated using an articulation test, that is, one with a fixed reading rate, it is found to be quite effective in interfering with the system. But when the person transmitting is permitted the freedom to vary his rate of speech, he can sometimes find a reading rate which will permit him to get through the babble and render the interference ineffective. Almost everyone is familiar with this problem from his experience in crowds, at cocktail parties, and in night clubs. Despite the hubbub, one can frequently find a code (a speed of delivery, a volume, or a tone of voice) which will permit conversation. Despite this ability, it is not possible, of course, to remove equivocation completely in finite time, but it is certainly possible to optimize to get the best information per unit time while accepting some error.

The second problem in coding procedure—the need for a general rule by which the particular code for use in a given situation can be determined—is illustrated by the data plotted in Figure 8-3. Once again the reader is asked to accept the figure and the calculations that have gone into it without further development. In this figure, the information rate per symbol or per message is plotted as a function of the message ensemble size wherein the signal-to-noise energy ratio per symbol is fixed. In this particular illustration, ensemble size is roughly analogous to vocabulary size in an articulation test, and each curve represents a different value of $2E/N_0$, a ratio which is so important to the understanding of both this and the radar problems that some explanation of it must be introduced here.

Stated briefly, the ratio $2E/N_0$ (where E is the signal energy and N_0 is the noise power per unit bandwidth) describes, for the case in which the signal is known exactly, the separation between the means of two statistical distributions divided by the standard deviation when these two distributions are along a unidimensional decision axis. One of these distributions is for the probability density of a measure on the decision axis when noise alone exists, whereas the second is conditional upon the existence of signal plus noise. It is further known that for small values of $2E/N_0$, one gets a large amount of equivocation with large ensembles, and a relatively small amount with small ensembles. As the value of $2E/N_0$ is increased, however, the equivocation in bits is decreased much faster in large ensembles than in

Rate $R = P \log_e M + (1 - P) \log_e \left(\frac{M}{M-1} \right) - H(p)$, M signals, MCPC (Articulation test)
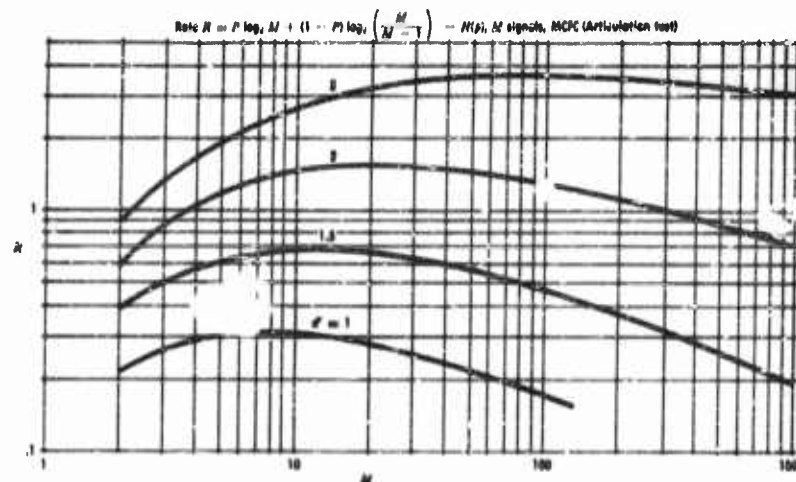
FIGURE 8-3   Rate of Information per Second as a Function of the Ensemble Size with Signal Energy Constant

small ones. (See Figure 8-3 for an illustration of this relation.) For each value of $2E/N_0$, therefore, there is an information rate per symbol or per message for each ensemble size, and for each such value, there is an optimum ensemble size for the rate of information flow.

The curves in Figure 8-3 illustrate this fact As mentioned above, each plots a different value of $2E/N_0$. According to the curves, for each signal-noise energy ratio, there is an ensemble size which leads to a maximum rate of information flow. Ensembles smaller than this do not have sufficient entropy to justify the capacity of the channel. In other words, the capacity of the channel is so great in relation to the entropy of the ensemble that a large amount of channel capacity is wasted. Conversely, ensembles greater than this are too large for the capacity of the channel. And as Shannon has shown if one attempts to transmit at a rate greater than the capacity of the channel, one incurs too great an increase in error.

One is tempted, of course, to think further in terms of Shannon's fundamental theorem and to search for the ideal code, one which will permit information flow at a rate near the channel capacity with arbitrarily small error. The theorem, however, assumes infinite time for transmission and cannot be applied to a finite system. It is based, moreover, on an existence proof. Although we know that the code exists, it is not defined. Yet despite these limitations, the concept remains useful, for it does state a bound. or standard, against which to compare performance. It represents an ideal in terms of which the efficiency of the real can be judged.

In voice communication studies, for example, one may find that through

Intelligent experimentation it is possible to determine general rules useful
in establishing the best of a set of procedures. Consider, as an illustration,
the reading rate established in a Standing Operating Procedure. It is desired
to determine the optimum reading rate based upon the actual situation
encountered. In other words, since the human component adds flexibility to
the communications system in that it allows the system to adapt to unfore-
seen circumstances, it would be valuable to establish rules upon which
individual standing operating procedures could be based. Any given pro-
cedure, therefore, would not be standard for all situations, but for particular
kinds of situations. Such knowledge should help increase the information
flow and decrease equivocation. It should enable one to approach the best
possible conditions.

8.2.2 Decoding

A second general area in which further study is essential is that of de-
coding and, most especially, the function of the human being on the receiv-
ing end of a communications system. To understand the problems raised
here, one must be somewhat familiar with not only the concepts of Shan-
non's theory, but also with three concepts from statistical decision theory.
Let us first state them flatly and then proceed to an explanation of each.
They are: (1) Woodward and Davies' assertion that a statement of *a
posteriori* probabilities contains all the information in the received signals;
(2) Van Meter's recently advanced conclusion that decisions on parts of a
message should be made to preserve as much information as possible, while
decisions on the total message should be made in relation to some other
criterion, such as the necessity for action; and (3) Wald's concept of
sequential observation as more efficient than fixed time observation in that
the same error restrictions can be satisfied with less average time. In the
following discussion, each of these ideas will be applied to the analysis of
the problem.

Basic to the discussion is Woodward and Davies' concept. They have
shown that a statement of *a posteriori* probabilities contains all the infor-
mation contained in a receiver input. This concept merely restates the one
made earlier—that it does not matter which end of the communication
system one looks through; the information remains the same. With decoding,
however, there is the danger of losing information. Doubtful signals must
be decoded with one meaning or another. As long as one is dealing only
with the signal, there is a possibility of correction of error. Once the decision
as to its meaning is made, however, information can and will be lost.

This concept is important because it yields a general rule for the question
of when to decode. According to this concept, one should decode only when

a decision is necessary for some consideration beyond those involving information. It may be necessary, for example, for the information to be relayed to another station by different means. The relay channel may not have the capacity to transmit the received signal. In such a case, the information must be decoded and sent. Or the nature of the message may be such—the presence of unidentified aircraft approaching in force—that one willingly incurs the loss of information so that necessary action may be taken. In the latter case, the need for action quite obviously overrides the need to preserve information. In other cases, however, where no such decisions are necessary, the preservation of information is of main concern.

This conclusion is certainly consistent with the concept recently advanced by Van Meter. He presents the view that all efforts should be directed toward preserving information until the need for action arises. It may be necessary, for example, to make a decoding decision because of a storage limitation. If this necessity should arise, the decision should be based upon criteria designed to preserve as much information as possible. To preserve information should be the major concern in every situation except that in which action becomes necessary. Only then should the criteria be changed to base the decision on getting the best possible results for the action. In other words, until there are uses for the information, all efforts should be directed toward storage of information as such. When the uses exist, the efforts should be directed toward optimizing the use.

The importance of these concepts is apparent when a human component on the receiving end of the system is considered. There are many things about him that one needs to know, most especially, how he stores and decodes the message. Does he decode unit by unit, symbol by symbol? In other words, is he like a novice trying to hold a conversation in a foreign language, translating each word as it comes up? Or does he, like a fluent speaker, preserve sentence after sentence in his mind, until he must act, that is, reply to the one he is speaking to? Does he, therefore, store information as far as possible, making decisions only when there is some compelling reason to do so? Suppose, further, that the time has come to act, that it is time, for example, for him to transmit to another station that information which he has so far received. Upon what criterion has he acted? Did he make his decision, as suggested by Van Meter, with a view toward preserving as much information as possible? If, on the other hand, he has decoded so that action may be taken, has the decision been based upon principles that will bring about the best possible outcome of the action? These questions clearly indicate that the inclusion of a human component in a communications system raises many problems which require further study.

Still another problem deserving further research is suggested by Wald's concept that a sequential observation is more efficient than a fixed-time observation in that the same error restrictions can be su.. ified in an average of half the time. The following illustration may clarify {   concept. Suppose a man has been invited to deliver a lecture. If he is important enough, he might conceivably elect to put his lecture on tape and have it played for his audience. To do this, however, he will have to estimate in advance the level of knowledge of his audience, the conditions in the auditorium, their attentiveness, and many other factors. He will have to make his recording on the basis of his estimate of what the average conditions will probably be. If he speaks in person, however, he can base some of his decisions on information he collects as he speaks. He may be able to judge from the facial expressions, nods of the head, etc., when each successive idea has been understood. Thus, he dwells on each idea until he receives confirmation from the audience that it is satisfied. On some of the ideas he may spend a longer time than he would have had he taped his speech in advance. In others he may find he does not need to spend as much time. On the average, assuming that it is a long speech containing many ideas, the speech would require less time per idea if he delivered the talk in person than if he taped it in order to achieve the same level of understanding on the part of the audience.

In effect, the tape-recorder example is analogous to the problem involving fixed-time observation. With a pre-established code—one constructed in advance—we must work upon the principle of averages. We accept a time unit which, on the average, should give us a reasonable probability of certainty, though we recognize and accept the fact that error is present. It is like planning a stock-buying campaign in advance. We know that we will make errors somewhere in our estimates and calculations, so we must make our plans on the average if we are to have any success. We have to accept error. In a similar fashion, with fixed-time observation in a finite system, we must also accept a degree of error.

On the other hand, the sequential observation—and our speaker permits his audience to act sequentially—enables us to operate more efficiently. Instead of operating on averages alone, averages which may not be at all descriptive of a given situation, the observer makes use of information immediately available. Thus, instead of observing for a fixed length of time, he observes only up to the point at which he achieves a certain level of confidence. To follow up our stock market analogy, with sequential observation, we would not plan the campaign in advance, but would make our decisions on purchases and sales on the basis of day to day information. We would have a clearer picture of the situation in which we are operating and thus should be able to reduce error.

Perhaps the most significant fact about sequential observation is that the observer has control of the length of observation time. Once he has the information, or has achieved his degree of confidence, he does not have to waste time by building in redundancy or error-correcting devices. He can accept the signal and go on to receive additional information. On the average, the observer who operates according to Wald's concept cuts the length of time to the point of removal of uncertainty to about half. Thus, he raises the degree of certainty per unit time.

Since sequential observation is so significant, one may well ask how this kind of process may be built into a communications system. The block diagram in Figure 8-1 contains the means which makes this kind of observation possible—the feedback channel which serves the same function as the audience's reaction did in the speaker example. Only a low-capacity channel is needed since its sole function is to transmit to the source the information that the receiver has now completed an observation. The receiver listens to the signal until he is in a position to accept one of the symbols with a satisfactory degree of confidence. At this point, he transmits over the feedback channel a single binary digit which states that he is now ready to start the observation of the next symbol. The capacity required of the feedback channel in this way is at most one bit per symbol. By using this way the human observer in a communications system can observe sequentially. He can satisfy the same error restrictions as in fixed-time observation in an average of half the time.

Since sequential observation might increase the efficiency of a communications system to such a marked degree, and since, also, the human component may serve as such an observer, it is clearly important to know the extent to which the human operator can act as a sequential observer. It will be noted also that this type of observation adds a desirable kind of flexibility to the communications system since it allows design decisions to be delayed until the precise environment is known in which the system is operating. Studies in this area, therefore, should provide information for improving the efficiency of communications systems.

8.2.3 Use of Redundancy

A third problem requiring further analysis and research is that of the use of redundancy in a communications system. The term *redundancy* as used here is *that fraction of the structure of the message determined not by the free choice of the sender, but rather by the accepted statistical rules governing the use of the symbols in question.* Varying degrees of redundancy may be built into any system. The amount of redundancy (actually about 50 percent) in the use of letters in the English language was already shown in the introduction, and it was discussed above (in terms of the second

concept from Shannon's theory) how redundancy can be built directly into a message. Such redundancy, of course, decreases the information transmitted but adds measurably to the degree of certainty with which the message is received. It serves, in effect, as an error-correcting device.

Fundamental to this discussion of redundancy is the first concept from Shannon's theory (that the rate of information flow is the same viewed from the standpoint of either the transmitter or receiver). This concept is very important in understanding the effectiveness of the performance of the total system. The reader should refer back to the first example given in the discussion of this concept, that in which the uncertainty of the transmitter was reduced from two bits to one bit, that of the receiver from one bit to none, so that in each case the uncertainty was reduced by one bit. In this case, the *a priori* uncertainty of the receiver was not matched to that of the transmitter. With the mismatch, there must be some equivocation at one end of the channel. It is only when the two ensembles are matched that equivocation can go to zero at both ends of the channel. Thus, the uncertainty of the receiver should be matched to the uncertainty of the transmitter, a matching which can be achieved only through the use of *a priori* information—that is, when the receiver has knowledge of the message ensemble.

This fact suggests the necessity of matching the ensemble of the transmitter with the ensemble of the receiver. Consider the following illustrations. Suppose the ensembles are completely unmatched and that an audience which understands only English suddenly finds itself being lectured to in Japanese. Since there is almost no relation between the languages, the message ensemble of the speaker is totally unknown to the listeners, and little or no communication takes place. The problem exists, moreover, even when the ensembles are partially matched. Suppose our lecturer speaks English, but has in mind a frame of reference for his words different from that of his audience (he is of a different religion or political party, or has a different social or economic view). It is possible that they may understand the words, but may be puzzled or confused as to his meaning. He is using an ensemble that they do not understand, and his message, perhaps, does not get through to them at all. Everyone is familiar with such common misunderstandings, even in the conversation of friends of long standing.

Essentially, therefore, the problem is one of the use of *a priori* information, of giving the receiver knowledge of the message ensemble. In our latter example, the speaker may evaluate his audience in advance and adjust his ensemble to the average to be expected. Or in seeing the puzzled looks on the listener's faces, he may restate his ideas in other forms—in other words, he may change the ensemble to fit theirs and so get through to them.

In still other instances, the audience itself may have acquired information concerning the point of view or frame of reference of the speaker. (Students frequently "case" their professors in this way.) Thus, they know in effect the message ensemble, and misunderstanding is less likely to occur. In a communications system, however, if the message ensemble is of considerable size, the memory of the human operator may be severely taxed.

It is important, therefore, to find means by which the human memory may be helped. To be sure, Miller has reported some interesting studies suggesting that human memory is quite restricted. Nevertheless, even though this may be true, there are means available which can certainly help. The human being does not have to rely upon unaided memory, for the use of memory aids can help him to a rather substantial and quite accurate memory. One such memory aid is the map used in communication studies at The University of Michigan.

This map is useful because it gives the receiver *a priori* knowledge of the message ensemble and thus, at least under test conditions, accomplishes the matching of the uncertainty of the receiver to the uncertainty of the transmitter. In this test, each of the messages is a route on the map illus-

FIGURE 8-4   Map Used as Visual Aid
to Redundancy

trated in Figure 8-4. The starting point is the position Y, heavily marked in the center of the map. Progress along the route is only to connected neighboring towns, and the route does not double back upon itself. Thus, there are four possible moves from V—to O, J, A, or C. If the move is to O, three possibilities follow: to E, R, or H. Each message consists of six towns, as for example, that marked, $OETRVE$.

The receiver may act either upon fixed-time observation or sequential observation. If he observes sequentially, the name of the town is repeated until the receiver transmits over the feedback channel a single digit denoting that he has received the signal with a satisfactory degree of confidence. By having the map available as a memory aid, the receiver can readily identify the received message as one of 972 possible routes. It is possible, however, that the received map may not be one of these routes. If it is not, it is still possible for the observer to choose the most likely of those messages actually in the ensemble. Thus, in the typical route used above, he would probably select that one as the correct

message, even though he actually received *ODTRVE*, a route which is not in the ensemble.

Results of this test suggest that memory aids of this type can complement human memory. Such use of redundancy, of giving the receiver knowledge of the message ensemble and thus matching his uncertainty to that of the transmitter, can increase the efficiency of communications systems. Further research in this area is, therefore, clearly indicated.

8.2.4 Summary

Analysis of the communications problem reveals a number of areas where further research is essential if we are to understand fully the nature of the system and the relation of the various parts to the efficient functioning of the whole. The role of the human being, whether he is on the transmitting or receiving end, is particularly significant here. As observed throughout, the human operator adds flexibility to a system. His presence may allow a number of parameters to remain unspecified until the system begins to operate in a particular environment. Once in the environment, the human being may adjust to the specific conditions he finds and perform more efficiently than present mechanical devices. Since he can serve so important a function, it becomes imperative to determine the effect of his behavior on the system as a whole.

From the standpoint of coding, one would like to know the effect of reading rate, volume, tone, inflections, and so forth on the efficiency of the system with a view toward determining general rules for establishing the best possible procedures for situations in which the operator may find himself. At the other end of the system, more information is needed on the capacity of the human component for storing or decoding information and on his ability to make decisions to optimize either information or utilities. We also need to know much more concerning the human being's ability to observe sequentially and to make use of redundancy in a communications system. We would like to know the extent to which he can incorporate *a priori* information. Increased knowledge in all of these areas will improve our understanding of the communications system. Such an understanding will then help in evaluating the effectiveness of any countermeasures program.

8.3 The Radar Problem

Although at first glance, the problem of the human component in a radar system would seem to be different from that in a communication system, it is in many ways much the same. To be sure, there are significant differences between radar and communication systems, differences which require that

they be treated separately. Nonetheless, as the present analysis will show, insofar as the human role is considered, the similarities are greater than the differences. This analysis is based upon a specific problem: the possible influence of human performance on the range of a radar. Despite the specific nature of the problem, however, it reveals certain pertinent questions regarding human performance in a radar system in general—questions which are the same as those encountered in the discussion of the communications system.

Before the analysis can be started, however, certain background material must be clearly understood. Since a radar system is a sensory system, the function of such a system is particularly pertinent here. Consider the block diagram shown in Figure 8-5. The area within the dotted lines on the
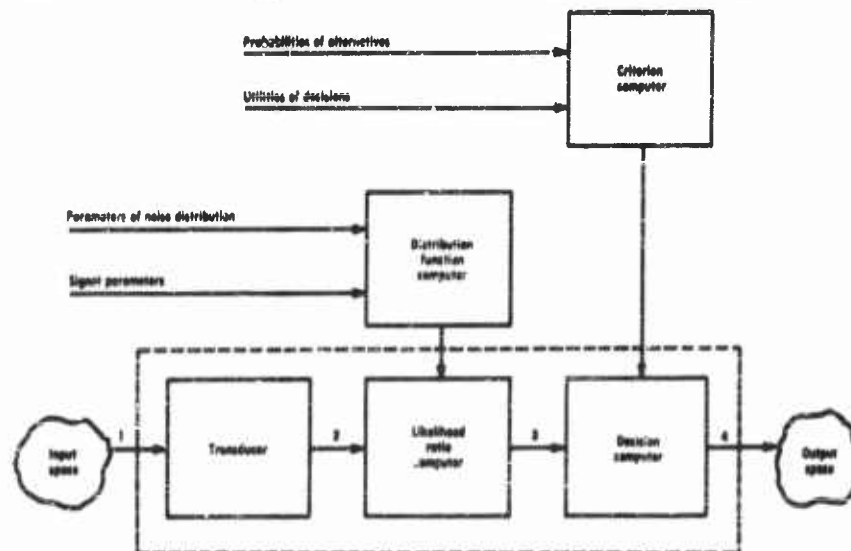
FIGURE 8-5   Relations of Functional Blocks in a Large System

chart is of concern. The sensory system is what gives one the observation; the $l(x)$ (likelihood ratio) is what one gets from the observation. The $l(x)$ computer calculates likelihood ratios and can compute only a finite number of them.

As can be seen from this chart, the sensory system is really a part of a larger one, and its efficient operation must depend upon instructions from the larger system. That is, the larger system must determine the likelihood ratios relevant to the particular situation and time. These instructions are sent into the sensory system by the distribution computer.

In similar fashion, before the question of decisions can be considered, it is necessary to know how the sensory system fits into the larger unit. For example, if the function of the sensory system is to transmit data to a data processing center, it should be concerned with transmitting as much information as possible. Its function should be to preserve information, not to act upon utilities. If, on the other hand, the output is used to make a decision to take action, the decision computer must consider the utilities of the decision. This problem, it will be observed, is similar to that discussed in Section 5.2.2. Finally, if the equipment is to be used for making decisions, the information fed into it from other parts of the system should be chosen by criteria for information preservation. That is, we should need the decision computer input information, not decisions.

The importance of this background becomes immediately apparent when it is applied to the particular problem of the radar system. The more information fed into a sensory system from the outside, the less has to be processed by the system itself. This fact is of the utmost importance in a radar system. A signal in noise has a capacity for carrying information, or, in other words, reducing uncertainty. Since this is true, the detection range of a radar is a function not only of the energy of the pulse return and the noise level at the receiver input, but also of the amount of information which must be processed. Thus, there is a relation between the amount of information to be processed and the range of the set.

If the *a priori* uncertainty is large, the energy of the return must be correspondingly larger to reduce the uncertainty to a designated value. If, however, *a priori* information has reduced the uncertainty, the energy required is less by an amount dependent upon the degree of the uncertainty. Consider, for example, Figure 8-6. Here it can be seen that it is possible to achieve a degree of certainty of, say, .98 with less energy as the number of alternatives among which the choice is made is decreased. It is clearly important, therefore, to reduce the uncertainty as much as possible before the set is asked to operate.

Suppose, for example, that our knowledge of the enemy is sufficient to ascertain that we can expect perhaps four possible targets, some 260 locations, and an infinite number of times during which the target may appear. Clearly, this is a vast amount of information, of uncertainty, to be processed. If, however, we can feed information into the set or system which will assign a high *a priori* probability to one particular target, a smaller set of locations, and a limited number of times, we decrease the uncertainty, require less energy to reach a designated level of certainty, and, in effect, increase the range of the radar. In other words, it would be possible to pick up the target while it is still farther out.
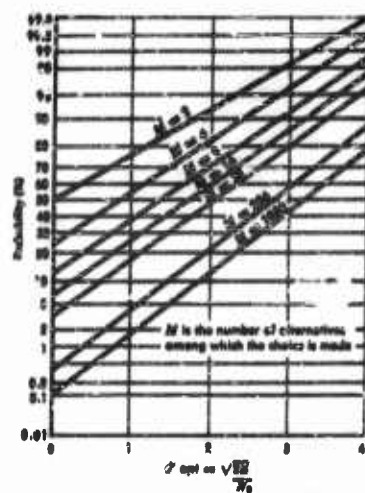
FIGURE 8-6 Probability of a Correct Choice as a Function of $2E/N_0$ with the Number of Alternatives as a Parameter

There are a number of ways, of course, in which this can be accomplished. A better definition of the signals that appear will certainly help. Also, when the radar is integrated into a system in which information from other sets is available, such information can be incorporated as *a priori* information for the radar operator. Thus, the sensory system (the radar) must be seen in relation to a larger system; the efficiency of the sensory system will depend upon instruction from that system. Only then can one know what the system can be expected to do.

Detection, moreover, should be based, as far as is possible, on sequential observation. That is, detection should come as a result of integrated information rather than as a result of a series of individual decisions. The presence or absence of a target should not be decided at each scan of the antenna for each resolvable unit of space as displayed on the radar scope. Rather, energy should be integrated from scan to scan until it is possible to make a decision in accordance with a certain level of confidence. It is important, therefore, that information be preserved until such time as a decision must be made. Preservation of information is the most important criterion until a decision for action must be taken.

To a large extent, a consideration of all these facts affects estimation of radar range only in the way in which the performance of the decision-making device is incorporated. It is independent of the physical or electrical properties of the radar except insofar as these may be modified in the light of information being collected. The flexibility introduced with the human component is important here. Under certain conditions, the operator may elect to employ a restricted or selected portion of scan or restricted range, operations which can influence the total energy of returns from targets. Thus, in a radar net, another set may give information concerning a certain target, including place, direction, and speed. Such information may lead the operator to elect a particular scan or range. He will require less energy to pick up the target and hence increase the range.

This section of the chapter, therefore, will consider the important question of the effect of operator performance on radar range—an effect which will

be considered in terms of fixed scan conditions, and fixed conditions for range of search. In terms of these assumptions, the evaluation of the signal and of receiver efficiency will be considered first. In this way, it should then be possible to determine operator efficiency. Such analysis will illustrate the relevant psychophysical areas for the study of radar operator performance.

8.5.1 The Assumed Signal Evaluation

Since operator efficiency must be seen in terms of both the efficiency of the signal and that of the receiver, these two areas must be considered

FIGURE 8-7   Three Points of Efficiency Measurements in a Radar System

first. The block diagram in Figure 8-7 will perhaps help clarify the problem. Here, the sensory system described in the introduction to this section is incorporated into a radar system, and the three points of measurement for signal, receiver, and observer efficiency are illustrated. In other words, a measure for determining signal efficiency is needed first. This factor must be modified by receiver efficiency, and the combined factor further modified by operator efficiency. In this way, one can learn what performance may reasonably be expected of a radar system, for the η (efficiency) is a variable that can be dealt with practically and is measurable in some experimental situations. There is no reason why it cannot be used in the future in increasingly complex ones.

Let us turn, then, to the question of signal evaluation to determine the basic factor in terms of which the receiver and operator efficiency must be seen. According to the theory of signal detectability, the most important parameter for such evaluation is the ratio $2E/N_0$, where E is the signal energy and N_0 is the noise power per unit bandwidth (see Appendix). This

ratio can be used to furnish the basis not only for decision tasks, but also for recognition ones. That is, when two possible signals are presented and one is subtracted from the other, the difference signal obtained has energy which goes into the $2E/N_0$ ratio, which tells how well one can discriminate between the two signals. Since this is true, it is therefore assumed that the signal evaluation at least furnishes the basis for calculating an expected value of the ratio as a function of the range of the radar from the target. Once the value of $2E/N_0$ has been determined, the efficiency of the human component can be worked out.

For the purpose of this discussion, a number of assumptions have been made. Figure 8-8 assumes a value of $2E/N_0 = 1.00$ when the range is 1.00. It further assumes that the signal power of the echo varies directly with the fourth power of the reciprocal of the range, that the number of pulses returned is a constant independent of range, and that the noise at the receiver input is independent of range. No effort is made to justify these assumptions; they have been made only to furnish a basis for the demonstration which follows. The necessary evaluative procedure is not dependent upon them. One could as well assume that the noise power is increasing with range, as might be the case if countermeasures were employed.

The curves shown in Figure 8-8 have taken this particular form because of the nature of the hypothetical experiment upon which they are based. The situation involves two fast moving objects coming together. The lower of the two curves shows the value of $2E/N_0$ expected for each target interception. The second is the integrated value assuming that the target is intercepted once every time the range is reduced by one tenth of its original value. Because of the particular nature of the experiment, the last value every time represents the major portion of the energy of the integration, the sum of the energy to that point. If the target is intercepted more frequently, then the integrated value is increased proportionately to the frequency of interception.
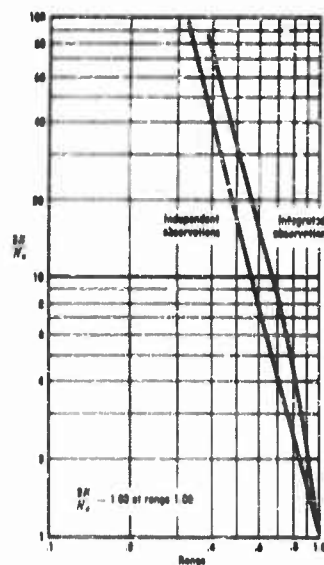
FIGURE 8-8 $2E/N_0$ as a Function of Range

Thus, if *a priori* information can be used to reduce the sector scanned, increasing the frequency of the intercept, the integrated value of $2E/N_0$ is expected to increase. With a more limited sector, one can sweep faster and get more interceptions; at the same time one picks up energy as well. The use of *a priori* information, therefore, limits the amount of information to be processed, increases the energy from the target return, and thus affects the range of the radar. Hence, from the point of view of the human operator, it is important to know the extent to which he can incorporate *a priori* information. Since the integrated values, moreover, imply sequential observation, the questions of the capacity of human memory and of the use of memory aids become important. The signal evaluation in terms of $2E/N_0$, therefore, raises significant psychophysical questions for further study.

These questions are pertinent, even though the assumptions underlying the particular values of $2E/N_0$ in Figure 8-8 are completely arbitrary. The necessary evaluative statement remains unchanged. The curve could be established on any set of realistic assumptions, leading, of course, to a more realistic curve. With realistic rather than assumed values of $2E/N_0$, one should be able to compute more accurately the efficiency of the operator.

8.3.2 Receiver Efficiency

Given the value of $2E/N_0$, derived from physical data, we must next modify it by a factor of receiver efficiency. This occurs at the second point of measurement shown in the block diagram in Figure 8-7. To a large extent, the functioning of this receiver and its efficiency will depend upon the way it is incorporated into the system and the function it is designed to serve. If there is an operator, as there is in the block diagram, the purpose of the receiver is only to transmit information, and its efficiency will depend upon the amount of information it can transmit in terms of its capacity.

If, on the other hand, there is no operator, the receiver must perform its functions, and it should act like the sensory system incorporated in the block diagram in the position of the operator. This, however, is an ideal observer; in practice, the efficiency would not be 1. Nonetheless, depending upon the function that the receiver must serve—to preserve information or to act upon utilities—one can assign an efficiency rating (η_R) which will modify the value of $2E/N_0$ derived from the evaluation of the equipment. Once again, η is a real value which can be used in the evaluation of the system.

8.3.3 Operator Efficiency

Once we have arrived at the value of $2E/N_0$ modified by η_R, we must next modify this combined value further by a consideration of an operator

factor, η_W. This occurs at the third point of measurement in the block diagram in Figure 8-7. For ease of calculation, it would be convenient to assume, as has sometimes been done, that this is a constant factor. In the past, for example, the human factor has been assigned an efficiency of .5. Experiments indicate, however, that this factor is not a constant, and that it only introduces unnecessary error.

In order to understand the variable factor η_H, designating the efficiency of the human operator, one must first understand the terms η, d', and the relationship between these factors and $2E/N_0$, since observer efficiency is related to these concepts. The definitions are based upon the block diagram
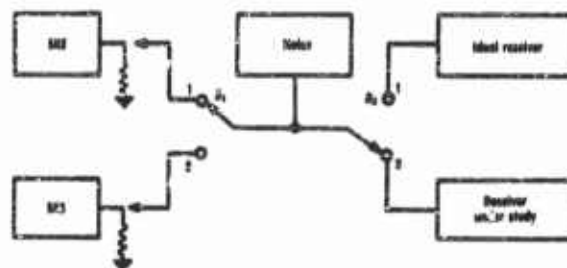
FIGURE 8-9 Block Diagram Conceptualizing Experimental Design (*from Tanner and Birdsall*)

shown in Figure 8-9, reproduced from Tanner and Birdsall. There are four channels the nature of which is determined by the position of the switches S_1 and S_2.

C_{11} is an ideal channel in which the transmitted signal is known exactly (SKE) and the receiver is ideal, that is, in relation to the particular transmitter. C_{12} is a channel in which the transmitted signal is known exactly and the receiver is one under study. C_{21} is one in which the signal is known statistically (SKS) and the receiver is ideal. (Note, however, that this is a different receiver from that in C_{11}, for, as the transmitter changes, the receiver must change, since it is ideal in relation to a specific transmitter). Finally, C_{22} is a channel with the signal known statistically and a receiver one under study.

Let us first consider an experiment with Channel C_{12}. S_1 is in position 1, S_2 in position 2. Signal energy E_{12} is employed, and noise power per unit bandwidth N_0 is added. The receiver's problem is to observe specified waveforms and to determine whether the waveform contains a signal plus noise or noise alone. Performance over a large number of times can be measured in terms of a detection rate $P_{SN}(A)$—the probability that if a signal was presented, it was accepted,—and a false alarm rate $P_N(A)$—

the probability of noise alone being accepted. The "hit rate" is therefore $P_{SN}(A)/P_N(A)$.

The second experiment is a mathematical calculation. It is similar to the first except that C_{11}, that is, an ideal receiver, is employed. With S_2 in position 1, therefore, the experiment is repeated, with the energy of the signal attenuated until the performance attained in the previous experiment is matched. This energy is E_{11}. The efficiency of the receiver under study for SKE (η_R), therefore, may be defined as E_{11}/E_{12}, where E_{11} is the energy level at which the performance is matched. The measure d' is then defined by the equation

$$(d')^2 = \eta_R \frac{2E_{12}}{N_0} = \frac{2E_{11}}{N_0}$$

That is, $(d')^2$ is the value of $2E/N_0$ necessary to lead to the observed performance, given SKE and its ideal receiver.

Let us next consider a second pair of experiments, both of which are mathematical calculations. The channel employed is C_{21}, that is, the signal is known statistically and the receiver is an ideal one for that statistical ensemble. Energy E_{21} and noise N_0 are employed, and a performance measure is established. S_1 is then moved to position 1, changing the channel to C_{11} (and, it should be noted, also changing the ideal receiver). Energy is attenuated until performance is matched at E_{11}. In the case of the SKS, this permits the calculation of the efficiency of the transmitter, η_t, for the SKS as E_{11}/E_{21}. At this point, it should be obvious that if one proceeds in this way with paired experiments, one can establish a receiver efficiency for any SKS, which might be expressed as $\eta_{R(SKS)} = E_{21}/E_{22}$.

One final definition remains to be made, η, that is, operator efficiency. Since $(d')^2$ is the value of $2E/N_0$ needed if one has ideal conditions, for a particular experiment $\eta = (d')^2/(2E/N_0)$, where E is the energy used in the experiment. The relationship, therefore, is one between the energy used under ideal conditions and that required in a real situation.

With these definitions, we can turn to the actual calculation of η_H, the efficiency of the human operator as it affects the performance of the system. In visual experiments, Tanner and Swets observed that d' for weak signals varied approximately as the square of signal intensity (energy). If the signal were known exactly, however, d' should vary as the square root of signal intensity. Since $(d')^2 = 2E/N_0$, this says that observer efficiency varies as a function of $(2E/N_0)^2$ for weak signals.

Although this information is important in the present context, the data acquired from the visual experiments are not in usable form, for in these

experiments, the values of $2E/N_0$ were not known precisely. In order to put these data into useable form for the purposes of this chapter, we must extrapolate from the results of auditory experiments. In these experiments, the values of $2E/N_0$ are known precisely and the η's can be calculated. Such extrapolation is, of course, very dangerous. It is done here only with the purpose of establishing an approximation. Quite obviously, further visual experiments and more precise information are necessary before one can be sure of the accuracy of the values of η.

The extrapolation is based upon the following theoretical observations: (1) signals known statistically, as they get large, approach in $(d')^2$ signals known exactly, and (2) observer efficiences in auditory experiments involving large signals approach values as high as .5 or greater. On the basis of these observations, it is assumed that when $2E/N_0 = 10$, the observer efficiency is .5 and becomes lower as $2E/N_0$ decreases, in accordance with the observation reported by Tanner and Swets. This is, it must be repeated, an extrapolation which should be checked experimentally. It may well be too optimistic.

These estimates are presented in Table 8-1. The η is derived as a result of the ratio $(d')^2/c/N_0)$ as defined above. The data from these columns are shown graphically in Figure 8-10, where the estimate of the human observer's performance is compared to the ideal. Once these values of η are accurately determined, one can proceed to modify the combined evaluation of the signal and receiver efficiency. The value of $2E/N_0$ as derived from signal evaluation and as modified by the efficiency of the receiver (η_R) would then be further modified by the operator efficiency factor (η_H) for that $2E/N_0$ as estimated in Table 8-1.

8.3.4 The Capacity of a Signal to Lead to a Correct Choice Among One of M Alternates

Now that the general process is understood by which the capacity of the

TABLE 8-1. ESTIMATION OF OPERATOR EFFICIENCY

$2E/N_0$	$(d')^2$	η
10	5	0.5000
9	4.1	.455
8	3.4	.425
7	2.8	.400
6	2.0	.333
5	1.4	.280
4	0.9	.225
3	0.53	.177
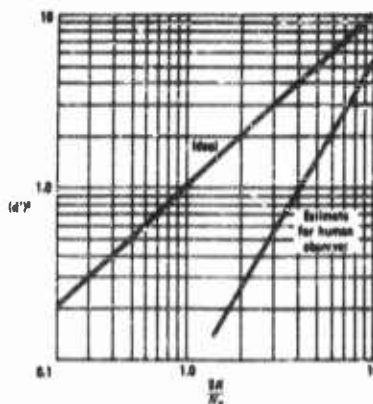
FIGURE 8-10 Estimate of Human Observer's Performance compared to the Ideal

system may be modified by factors of receiver and operator efficiency, we may return to the question of the evaluation of the equipment, as presented in Section 8.3.1 above, to show specifically how the efficiency may be affected by the use of *a priori* information.

In a short unpublished memo, Peterson and Birdsall have considered the problem of making a correct choice from M equally likely orthogonal signals. If the two statistical distributions basic to the calculation of $2E/N_0$ (see Appendix) are consulted, this is a case in which $M - 1$ observations are from the distribution conditional upon noise alone, and one observation is from the distribution conditional upon signal plus noise. The probability of a correct choice is the probability that the observation from signal plus noise is greater than the greatest of the $M - 1$ observations from noise alone. For the ideal case, this is given in the following equation:

$$P(c) = \int F^{M-1}(x) f\left(x - \frac{2E}{N_0}\right) dx$$

where

$$F(x) = \int_0^x f(x)\ dx$$

and $f(x)$ is the possibility density for the observation x.

Peterson and Birdsall have constructed a table (see Table 8-2) based on an approximation of this equation, and shown here graphically in Figure 8-6. From their approximations, Figure 8-11 has been constructed. This figure shows the value of $2E/N_0$ necessary to lead to a correct decision as a function of the *a priori* uncertainty ($\log_2 M$) with performance criteria of .85, .90, and .95 probabilities. As can be seen from the graph, the larger than *a priori* uncertainty, the greater the energy needed to reduce the uncertainty to a designated value.

8.3.5 Estimating M

It is assumed here that if the error of target location is normally distributed over a single dimension, then the uncertainty ($\log_2 M$) of target location is approximately the same as in the case of a target equally likely over the equivalent rectangular distribution. An approximation of M is then given by

$$M = \frac{\sigma \pi}{2\ \text{Antenna Beamwidth}}$$

where σ is the estimate of error in *a priori* estimate of target position.

TABLE 8-2. THE PROBABILITY OF A CORRECT CHOICE OF ONE OF M EQUALLY LIKELY ORTHOGONAL ALTERNATIVES (from Peterson and Birdsall)

M (d')	0	0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0
2	0.50	0.64	0.76		0.92	0.984			
3	.33	.48	.63		.87	.97			
4	.25	.39	.55		.83	.96			
8	.125	.24	.39		.71	.92	0.987	0.999	
16	.0625	.14	.26	.42	.60	.87	.975	.997	
32	.03125	.08	.17	.31	.48	.80	.957	.995	
256	.00391	.02	.05		.22	.57	.866	.979	0.9984
1000	.0010	.01	.02		.12	.41	.763	.951	.9950
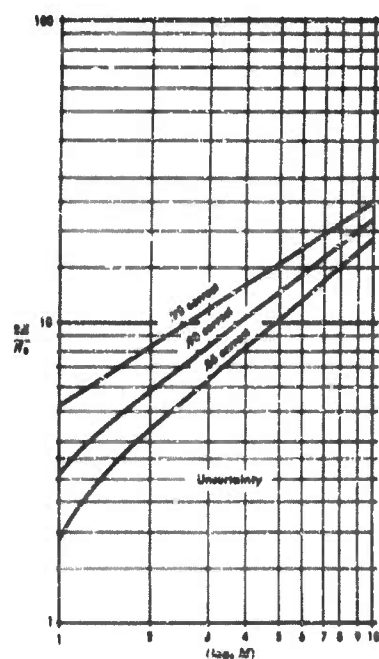
FIGURE 8-11 $2E/N_0$ Required to meet Criterion as a Function of Uncertainty

8.3.6 Summary

Although the radar problem is in many ways fundamentally different from the communications problem, the analysis of the radar range problem presented above illustrates certain fundamental similarities as far as the human component is concerned. The relevant psychophysical areas for the study of radar operator performance are not fundamentally different from those illustrated in the communications problem. In both, the relevant questions are the same. Just as the use of redundancy in the communications problem raised the question of the ability of the human being to incorporate *a priori* information, so does the radar problem, insofar as such information affects the energy requirements and, hence, the range of the system.

Thus, the areas in which further study is required may be summarized by means of the following questions: (1) Can the operator incorporate *a priori* information? (2) What is the extent of his memory, and what kind

of aids can be introduced to supplement this memory? (3) Can the operator act as a sequential observer? (4) Can the operator optimize information? These questions are very similar to those asked in the communication problem. All will require further study before one can be sure of just how the human component may be incorporated into a system and how his incorporation will affect the performance of the system as a whole.

8.4 The Countermeasures Problem

All of the questions treated in the preceding two sections are pertinent to the countermeasures problem. Most especially, the dominant idea of the foregoing discussion needs to be stressed again: the efficiency of the system depends to a great extent on the specific job the system is designed to perform; the more clearly the situation in which the system will act can be specified, the more accurately can its efficiency be evaluated. The role of each of the factors in the system must be carefully studied for two important reasons. First, we must clearly understand the effect each factor—including the human one—has on the system as a whole so that we can accurately evaluate the system. Second, once given this information, we must strive toward the formulation of a more general rule in terms of which the system may be evaluated.

The importance of this second concept cannot be overemphasized. Although the system must be studied in terms of specific situations, the data acquired may not be entirely useful in the evaluation of equipment if that equipment should be used for an entirely different purpose. Since the efficiency depends to a great extent on the particular situation or game, a change in that situation may render the data derived from the first situation all but useless. Clearly, what we must seek is a more general rule, a more general means of evaluation which can be used to cover a variety of specific situations. The data should be in some form—still to be determined—which will not be designed to tell the user the specific evaluation of the equipment, but which *will* enable him to calculate that efficiency for himself given the specific situation in which he is engaged. It is this second concept which is of major concern in this chapter.

As in the preceding two sections, the analysis of the countermeasures problem must proceed from a realization that the problem is derived from the fact that we are involved with a system. As such, it can best be understood in terms of the material that has already been covered. Consider first the relation to the communications problem. If we consult the block diagram in Figure 8-12, illustrating the basic countermeasures problem, we can see immediately that what we have here is really a modified version of Shannon's fundamental diagram of a communication channel (Figure 8-1), with the operator and jammer added. As one would expect, therefore, the
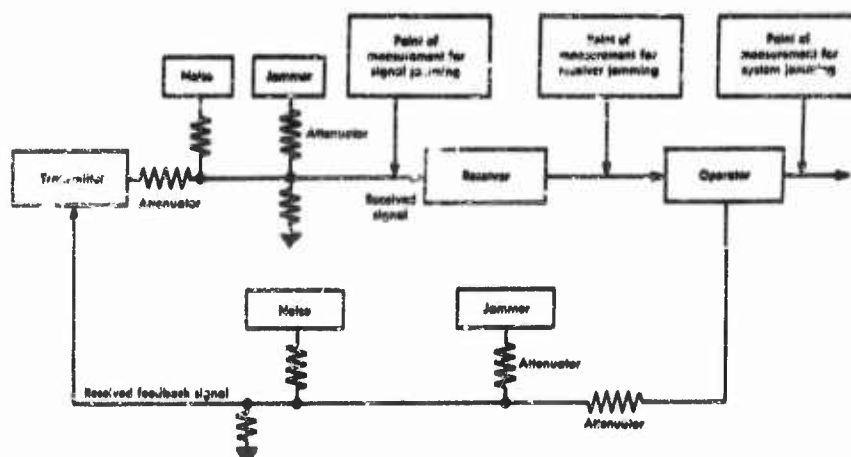
FIGURE 8-12 Block Diagram Illustrating Basic Countermeasures Problem

fundamental questions of information, channel capacity, entropy, and so forth, all apply here, especially insofar as they are affected by the presence of the jammer. In a similar fashion, the psychophysical questions raised by the incorporation of a human component are equally pertinent in this context.

Certain similarities will also be noted between this diagram and the one basic to the radar problem shown in Figure 8-7, wherein the sensory system was incorporated into the radar. Of the greatest importance are the three points of measurement used to determine the effectiveness of jamming. These correspond generally to the three points of measurement for determining equipment, receiver, and operator efficiency in the earlier problem and may be easily understood in terms of that problem. Thus, the question of determining the value of η, basic to the radar problem, is pertinent to the countermeasures problem as well, and has to be worked out for each factor in the system, including the human one. The diagram in Figure 8-12, therefore, is presented here to assist the reader, familiar with the preceding problems, in identifying the factors involved in evaluating the effectiveness of a jamming tactic.

The factors involved in such evaluation may be divided into three main categories, two of which relate to the communications system in general, and a third which is related to the countermeasures problem in particular. The first and second categories may be distinguished from each other according to their relation to the specific situation. The word "specific," however, is used here in a somewhat unusual sense. It does not refer to the particular physical parameters involved—such as distance or terrain—

but rather to those elements in the situation determined by the particular game being played. The first group of factors is specific to the game; the second is not.

Those factors, among others, which are specific to the particular situation may be listed as follows:

(1) Capacity required of the forward channel.
(2) Coding scheme employed.
(3) Decoding scheme employed, including such factors as the effectiveness of the operator.
(4) Use of the feedback channel.
(5) Criterion of acceptable performance, such as
    (a) Permissible error rate,
    (b) Permissible miss rate,
    (c) Importance of time.

All of these areas are familiar to the reader from the discussion of the comunications problem in Section 8.2 above, and a number of the problems concerning at least some of the areas have been touched upon there. They should require, therefore, no additional discussion here.

The physical factors involved—those not considered specific to the game—are as follows:

(1) Capacity of the forward channel.
(2) Distances involved.
(3) Atmospheric attenuation.
(4) Terrain factors.
(5) Condition of equipment.
(6) Regulation of power supply.
(7) Susceptibility of equipment, such as
    (a) Saturation,
    (b) Effect of nonlinear elements,
    (c) Detectability by jammer.

The first six of these factors are also important in the third category, that which, as noted, relates most particularly to the countermeasures problem. To this third category, two more factors can be added:

(1) The ability to take advantage of the detectability of the jamming target, and
(2) The ability to transmit an adequate jamming signal.

Once these three main sets of factors are understood, we may proceed to the question of the evaluation of a jamming operation.

As in the case of the radar problem, the effectiveness of the system cannot be measured at only one point, since to measure the effectiveness in this way fails to take into account the relative importance of each of those factors which contribute to the evaluation, and leaves the observer in the dark concerning the relation among those factors. For this reason, in the block diagram in Figure 8-12, three points of measurement are indicated—for signal jamming, receiver jamming, and observer jamming—points which do permit the isolation of the relevant factors and which should furnish a much sounder foundation for collecting information upon which to base future research and development programs. Let us examine each of these points in turn, therefore, to see the relation among them and the relative effect each has on the evaluation of the countermeasures system as a whole.

The first point of measurement comes before the signal enters the receiver and is the point for measuring signal jamming. As defined by Hok, this is actually a measure of the ability to reduce the capacity of the channel up to the input of the receiver. Remember that capacity is defined in terms of the information (in the special sense of communication theory) which the channel can transmit. The jamming signal reduces that capacity and increases the uncertainty with which the signal is received. The measure is made at this point rather than at the second point because any but an ideal receiver will further increase that uncertainty. An ideal receiver uses all the information at the input, and were such a one employed, then the measure at point one (for signal jamming) would be the same as that at point two (for receiver jamming). In an actual case, however, the receiver will add additional uncertainty, and we wish to isolate the first important factor. This factor is a measure of the additional entropy (the degree of randomness) of the input signal as a result of the jamming signal.

The second point of measurement is that for receiver jamming. What we wish to isolate here is the degree of uncertainty which comes as a result of the receiver, that is, of qualities inherent in the equipment itself. Measurement at this second point requires, therefore, a measure of the additional entropy at this point, and then the isolation of that part of the additional entropy which is due specifically to the receiver. Because realizable receivers have a finite range and nonlinear elements, the efficiency of the receiver may depend upon conditions at the input. Consequently, if the efficiency of the receiver changes as the signal is jammed, then the receiver is contributing to the entropy beyond that of signal jamming. Only with an ideal receiver are the measures at points one and two the same. To measure only at point two, therefore, will give a false estimate of the efficiency of signal jamming. Similarly, a measure at only point one would ignore the fact that realizable receivers add to the uncertainty with which the signal is received. Measures at both points, therefore, are required if we are to get an accurate estimate of the efficiency of the jamming tactic.

This concept will perhaps be further clarified if it is considered in terms of a specific example. Consider a channel, the capacity of which is measured in terms of bandwidth and the signal-noise ratio, where only noise relevant to the signal is considered in calculating the ratio. Designate this capacity as D. Now if the measure is made at the input of the receiver, that is, at our first point of measurement, the effectiveness of signal jamming can be expressed by the following ratio:

$$\frac{D_0 - D_j}{D_0}$$

where D_0 is the capacity of the channel without jamming and D_j is the capacity with jamming. One can learn thereby the degree by which the jamming signal has reduced the capacity of the channel. This then yields a measure of the effectiveness of the signal jamming.

Let us proceed to the second point of measurement, that for receiver jamming. If the measure is made here, the effectiveness can be expressed by the following ratio:

$$\frac{\eta_0 D_0 - \eta_j D_j}{\eta_0 D_0}$$

where η_0 is the efficiency of the receiver when the input signal is D_0 and η_j is its efficiency when the input signal is D_j. What we have done in effect is to modify the measure of the effectiveness of the signal jamming by a factor of receiver efficiency to arrive at an over-all estimate of the efficiency of the jamming operation to this point. This last expression can be rewritten as follows:

$$\left(1 - \frac{D_j}{D_0}\right) + \frac{D_j}{D_0}\left(1 - \frac{\eta_j}{\eta_0}\right)$$

where $[1 - (\eta_j/\eta_0)]$ can be considered the effectiveness of the receiver jamming. This rewriting of the formula is useful because it isolates the term $[1 - (\eta_j/\eta_0)]$. If one is interested in the susceptibility of the receiver, this is the factor which should be studied, because it expresses that part of the effectiveness of the jamming operation that comes as a result of qualities inherent in the receiving equipment itself.

Let us proceed further to the third point of measurement, that for system jamming. The third factor which enters here is the efficiency of the human

operator, a factor which, as was shown in the radar problem in Section 8.3, modifies further the efficiency of the system as a whole. To account for this third factor, let us extend the measure as it is rewritten in the paragraph just above. It can be extended to n terms. If it is extended to three terms to take in the factor of operator efficiency, the measure reads:

$$\left(1 - \frac{D_j}{D_0}\right) + \frac{D_j}{D_0}\left(1 - \frac{\eta_{RJ}}{\eta_{RO}}\right) + \frac{D_j\eta_{RJ}}{D_0\eta_{RO}}\left(1 - \frac{\eta_{HJ}}{\eta_{HO}}\right)$$

where the η's are distinguished by their subscripts as η_R, the efficiency of the receiver, and η_K, the efficiency of the human operator. Just as the second step in the measurement process isolated a term useful in studying the susceptibility of the equipment, this expression isolates the term $[1 - (\eta_{HJ}/\eta_{KG})]$, which can be considered the effectiveness of observer jamming. This would include the additional uncertainty, over and above that caused by signal and receiver jamming, which can be attributed to the human operator. This then is the factor which should be studied if one is interested in the susceptibility of the human being. It expresses that part of the jamming operation that comes about because of the nature of the human operator.

The total expression, which takes into account all three factors, describes the effectiveness of the system jamming. Each of the factors in parenthesis represents the effectiveness at one of the points of measurement: the signal jamming, the receiver jamming, and the observer jamming. Any measure which considers only the system as a whole fails to isolate these three factors which contribute to the specific effects in any single test. On the other hand, measures taken at each of three points specified in the block diagram in Figure 8-6 do permit their isolation. They provide a basis for studying individual factors, such as equipment evaluation and observer efficiency. They should furnish, therefore, a more solid foundation for the collection of information upon which future research and development programs may be based.

The form in which these data should be presented, however, is still to be determined. It seems obvious that in each stage of the measurement, the results of the measures are, to some extent at least, dependent upon the particular situation or game. The η's are certainly variable factors, their values varying with the specific situation. This specificity must be realized and studied, not only so that we can understand the role of each factor in a specific game, but more especially so that we can eventually determine more general means of evaluation, means which can be used for a variety of specific situations.

A concrete example will perhaps clarify the issue here. What we are seeking are data which can be presented in a way similar to that in which a manufacturer presents specifications on an oscilloscope. He does not present a number which evaluates the equipment in terms of a particular use the customer has for the product. Rather, he presents a set of data which permits the customer to calculate the value of the product for his own use. The data are designed to permit this calculation for a large number of uses. In like manner, we must find a means for presenting data in a way that will permit the evaluation of not just one jamming operation. Rather, the data should be in such forms as will permit those engaged in countermeasures programs to evaluate their own specific situations for themselves.

We should seek, therefore, a general rather than a specific rule. Once we are aware of this fact, we can state the fundamental problem. Laboratory data on jamming tests should be presented in this way: there should be a set of specifications permitting their use in evaluating a large number of specific situations, rather than a number evaluating a specific situation. What the specifications should be, and the ways in which one uses them to estimate the results in a specific situation, are closely related questions. The answers to them depend, however, on knowledge in areas which are still relatively undeveloped. As has been stressed throughout this chapter, the function of the human component in any system is one such area.

Particularly significant is an understanding of the way in which the human being introduces flexibility into the total system, since his ability to incorporate *a priori* information, to function as a sequential observer, or to optimize information or utilities can have a significant effect upon the functioning of any system. Furthermore, his self-evaluating and self-adjusting qualities may lead to such flexibility that he might even be able to offset, at least partially, the effect of jamming in a communications system. The relation of the human component to the system as a whole and his possible effect on its efficiency are, therefore, areas where considerable research is certainly warranted.

APPENDIX
Mathematical Calculations of the Discriminability of Two Signals*

The purpose of this appendix is to describe the mathematical development of the determination of the restrictions placed on performance in a detection experiment by the environmental conditions.

---

*The mathematical development presented here follows closely that of Peterson, Birdsall, and Fox.

A signal is defined here as a voltage waveform. In a well-defined time interval, one of two signals perturbed by noise is presented to an observer or a receiver. It is the observer's task to state which of the two signals was contained in the input waveform.

In this analysis, the voltage waveforms, the signals, are precisely defined functions of time, $S_1(t)$ and $S_2(t)$, entirely contained within the observation interval, t to $t + T$. The precise definition of the waveforms means that, if the waveform exists in the interval, the voltage amplitude at any instant in time, t_i, is a precisely specified value S_i.

The perturbing noise is assumed to be Fourier series bandlimited white Gaussian noise which is added to the signal. This set of assumptions is made to permit discrete statistical analysis. Since the noise is series band-limited, a noise waveform within the interval can be precisely specified by $2WT$ measures of voltage amplitude, where W is the bandwidth and T is the duration of the interval. The fact that the noise is white implies that the power density is uniform at every frequency within the band W. The Gaussian assumption requires the voltage amplitude of the noise to be a normal variate with mean zero and variance, or noise power, N. The assumption that the noise is added to the signal states that at each instant of time in the interval, the voltage of the input waveform $x(t)$ is the sum of two voltages, the signal voltage and the noise voltage.

The purpose of this appendix is to show how well an optimum receiver can do in the task of specifying which of the two waveforms $S_1(t)$ or $S_2(t)$ is contained in an input waveform $x(t)$ when the receiver is faced with the restrictions defined by the assumptions outlined above. The analysis treats the receiver as testing statistical hypotheses. Measures of performance are statements of averages expected of the receiver over an infinite sequence of independent observations.

It has been shown that an optimum receiver bases its decisions on a likelihood ratio criterion, Eq (8-5). That is to say, the optimum receiver accepts one of the hypotheses whenever the likelihood ratio is a value greater than a weighting function β, otherwise it accepts the other. If monotonic transformations of both likelihood ratio and the weighting function are incorporated, an equivalent decision rule can be developed. The task of the analysis is to study the distribution either of the likelihood ratio or some appropriate monotonic function of the likelihood ratio first under the condition that one of the signals $S_1(t)$ exists, and then under the conditions that $S_2(t)$ exists.

In order to perform the analysis, a sampling theorem based on the series bandlimited assumption is employed. The purpose of the sampling theorem is to permit the use of discrete statistics. It says essentially that the input waveform, the noise, and the signal can each be specified completely by

$2WT$ independent voltage amplitudes ($x(t) = x_1, x_2 \ldots x_{2WT}$; $S_1(t) = S_{1,1}, S_{1,2} \ldots S_{1i} \ldots S_{1,2WT}$; $n(t) = n_1, n_2 \ldots n_i \ldots n_{2WT}$). From this it follows that

$$\sum_{i=1}^{2WT} (S_{1i})^2 = 2W \int_{i}^{i+T} \overline{S_1(t)^2}\, dt = 2WE_1 \qquad (8\text{-}1)$$

where E_1 is the energy in the voltage waveform $S_1(t)$, assuming that the waveform exists over a one ohm resistance.

The likelihood ratio is defined as

$$l[x(t)] = \frac{f_{s_1(t)}[x(t)]}{f_{s_2(t)}[x(t)]} \qquad (8\text{-}2)$$

where $f_{s_1(t)}[x(t)]$ and $f_{s_2(t)}[x(t)]$ are probability densities conditional upon the waveform $x(t)$ resulting from $S_1(t)$ and $S_2(t)$ respectively. Since the $2WT$ points are independent and specify the waveform in its entirety,

$$f_{s_1(t)}[x(t)] = \prod_{i=1}^{2WT} f_{s_{1i}}(x_i)$$

and $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (8\text{-}3)$

$$f_{s_2(t)}[x(t)] = \prod_{i=1}^{2WT} f_{s_{2i}}(x_i)$$

Since $n_i(t)$ is a normal variate, x_i is a normal variate with mean S_{1i} or S_{2i} and variance N. Therefore,

$$f_{s_{1i}}(x) = \left(\frac{1}{2\pi N}\right)^{1/2} \exp\frac{-(x_i - S_{1i})^2}{2N}$$

and $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (8\text{-}4)$

$$f_{s_{2i}}(x) = \left(\frac{1}{2\pi N}\right)^{1/2} \exp\frac{-(x_i - S_{2i})^2}{2N}$$

Substituting Eq (8-4) in (8-3), and the result in (8-2) leads to the following equation:

$$l[x(t)] = \frac{\exp\left[-\sum(x_i - S_{1i})^2\right]/2N}{\exp\left[-\sum(x_i - S_{2i})^2\right]/2N} \tag{8-5}$$

It is now appropriate to consider the natural logarithm of the likelihood ratio:

$$\ln l[x(t)] = \frac{-\sum x_i^2 + 2\sum x_i S_{1i} - \sum S_{1i}^2 + \sum x_i^2 - 2\sum x_i S_{2i} + \sum S_{2i}^2}{2N}$$

$$\ln l[x(t)] = \frac{2\sum x_i S_{1i} - \sum S_{1i}^2 - 2\sum x_i S_{2i} + \sum S_{2i}^2}{2N} \tag{8-6}$$

Examination of Eq (8-6) shows that x_i is the only variable. Since at each of the i points x_i is a normal variate regardless of which signal is present, the $\ln l[x(t)]$ being a sum of independent normal variates is likewise a normal variate.

It remains only to determine the means and variances of the two distributions conditional upon the inclusion first of $S_1(t)$ in $x(t)$ and then of $S_2(t)$ in $x(t)$.

If $S_1(t)$ is included, then the expected value of each x_i is S_i. Substituting this in Eq (8-6) and letting M_1 be the mean of the $\ln l(x)$ conditional upon the inclusion of $S_1(t)$ leads to the result

$$M_1 = \frac{2\sum S_{1i}^2 - \sum S_{1i}^2 - 2\sum S_{1i} S_{2i} + \sum S_{2i}^2}{2N}$$

$$M_1 = \frac{\sum S_{1i}^2}{2N} + \frac{\sum S_{2i}^2}{2N} - \frac{\sum S_{1i} S_{2i}}{N} \tag{8-7}$$

Defining $N_0 = N/W$ and employing the sampling theorem (Eq 8-1), Eq (8-7) becomes

$$M_1 = \frac{E_1}{N_0} + \frac{E_2}{N_0} - \frac{\sum S_{1i} S_{2i}}{N} \tag{8-8}$$

The last term is a correlation term. Letting

$$\rho = \frac{\sum S_{1i} S_{2i}}{\sqrt{\sum S_{1i}^2}\,\sqrt{\sum S_{2i}^2}}$$

$$\frac{\sum S_{1i} S_{2i}}{N} = \sqrt{\frac{2WE_1}{N}}\,\sqrt{\frac{2WE_2}{N}}\,\rho$$

$$= \sqrt{\frac{2E_1}{N_0}} \sqrt{\frac{2E_2}{N_0}} \, \rho \tag{8-9}$$

Substituting Eq (8-9) into Eq (8-8):

$$M_1 = \frac{E_1}{N_0} + \frac{E_2}{N_0} - \rho \frac{2E_1}{N_c} \frac{2E_2}{N_0} \tag{8-10}$$

By examination of Eq (8-6), it can be seen that the mean M_2 of the distribution conditional upon the inclusion of $S_2(t)$ is the negative of M_1:

$$M_2 = - \frac{E_1}{N_0} - \frac{E_2}{N_0} + \rho \sqrt{\frac{2E_1}{N_0}} \sqrt{\frac{2E_2}{N_0}} \tag{8-11}$$

The difference between the means is

$$d = \frac{2E_1}{N_0} + \frac{2E_2}{N_0} - 2\rho \sqrt{\frac{2E_1}{N_0}} \sqrt{\frac{2E_2}{N_0}} \tag{8-12}$$

Now consider the variance σ_1^2 of Eq (8-6) conditional upon the existence of $S_1(t)$. The terms $\sum S_{1i}^2$ and $\sum S_{2i}^2$ are constants and consequently contribute no variance.

$$\sigma_1^2 = \frac{(2\sum x_i S_{1i} - 2\sum x_i S_{2i})^2}{4N^2} \tag{8-13}$$

Since the i points are independent and values of S_{1i} and S_{2i} are constant for each i, this expression can be rewritten as

$$\sigma_1^2 = \frac{\sum x_i^2 (\sum S_{1i} - \sum S_{2i})^2}{N^2} \tag{8-14}$$

The expected value of $\sum x_i^2$ is N by definition; therefore,

$$\sigma_1^2 = \frac{\sum S_{1i}^2}{N} + \frac{\sum S_{2i}^2}{N} - \frac{2\sum S_{1i} S_{2i}}{N}$$

Employing the sampling theorem:

$$\sigma_1^2 = \frac{2E_1}{N_0} + \frac{2E_2}{N_0} - 2\rho \sqrt{\frac{2E_1}{N_0}} \sqrt{\frac{2E_2}{N_0}} \tag{8-15}$$

It is obvious that

$$\sigma_2{}^2 \;=\; \sigma_1{}^2$$

Thus the two distributions are normal with equal variance. The value of d'_{opt}, for the optimum receiver is the difference in the means divided by the standard deviation:

$$d'_{opt} = \sqrt{\frac{2E_1}{N_0} \;+\; \frac{2E_2}{N_0} \;-\; 2\rho\sqrt{\frac{2E_1}{N_0}}\sqrt{\frac{2E_2}{N_0}}} \qquad (8\text{-}17)$$

It should be pointed out that ρ, a correlation term, describes the correlation between the mathematical description of the two signals. In the case considered in the text, the two signals are pulses of sine waves, differing only in amplitude. In this case $\rho = 1$. Therefore

$$d'_{opt} = \sqrt{\frac{2E_1}{N_0}} \;-\; \sqrt{\frac{2E_2}{N_0}} \;=\; \sqrt{\frac{2E_\Delta}{N_0}} \qquad (8\text{-}18)$$

where

$$E_\Delta \;=\; \int\limits_{t}^{t+T} [S_1(t) - S_2(t)]^2 \, dt$$

It is necessary to point out again that the result depends on the particular set of assumptions described at the beginning of the appendix. If the assumption of series bandlimit had been different, the result would have been different. For example, if the noise is assumed to be transform bandlimited and therefore analytic, Sleppian has shown that the signal is perfectly detectable since in this case the noise is deterministic from $-\infty$ to $+\infty$. Proofs of perfect detectability require mathematically precise measurement. Any error of measurement, no matter how small, invalidates the proof. So far the assumptions employed indicate that if the signal is finitely detectable, the series bandlimit assumption at least leads to a result nearly that of any other set of assumptions so far examined. The Fourier series bandlimited assumption is therefore accepted as adequate for the present.

BIBLIOGRAPHY

1. Shannon, C. E., and W. Weaver, "The Mathematical Theory of Communication," University of Illinois Press, Urbana, 1949.

2. Woodward, F. M., and T. L. Davies, Information Theory and Inverse Probability in Telecommunications, *Proc. IEE*, Vol. 99, Part III, pp. 37-44, March, 1952.

3. Van Meter, D., Talk at IRE Convention, 1956.

4. Wald, A., "Sequential Analysis," John Wiley & Sons, Inc., New York, 1947.

5. Miller, G. A., Human Memory and Storage of Information, *Trans. IRE*, PGIT-2, 1956.

6. Peterson, W. W., T. G. Birdsall and W. C. Fox, The Theory of Signal Detectability, *Trans. IRE*, PGIT-4, September, 1954.

7. Van Meter, D., and D. Middleton, Modern Statistical Approaches to Reception in Communication Theory, *Trans. IRE*, PGIT-4, 1954.

8. Tanner, W. P., Jr., Theory of Recognition, *J. Acoust. Soc. Am.*, Vol. 28, pp. 882-888, 1956.

9. Tanner, W. P., Jr., and J. Swets, Human Use of Information. 1. Signal Detection for the Case of the Signal Known Exactly, *Trans. IRE*, PGIT-4, 1954.

10. Peterson, W. W., and T. G. Birdsall, The Probability of a Correct Decision in a Forced Choice Among M Alternatives (unpublished memo).

11. Hok, G., "A Study of Optimum Jamming Signals," Electronic Defense Group, Technical Report No. 17, University of Michigan, October, 1954. (SECRET)

9

The Intercept Receiver

W. R. RAMBO

The intercept receiver takes many useful forms. A large receiver of elaborate design may constitute the primary unit of an electronic intelligence (ELINT) intercept system; as such, it must operate in an integrated manner with antennas and data handling units of appropriate complexity. A second receiver, one featuring small size, light weight, and simplicity, might be a primary alerting device in bomber defense. A third receiver might be so completely integrated physically and electrically into a jamming system as to be scarcely recognizable as a receiver. A fourth might be designed to fit conveniently into a brief case. The keynote here is variety—in operational use, in signal environment, in the physical requirements imposed on the receiver, and, consequently, in the practical forms of receivers currently in use. Present technology does not permit a "universal intercept receiver." There is no single basic receiver technique that is close to being optimum for all uses (see Figure 9-1).

Despite the important variations, intercept receivers commonly employ standard basic receiver circuits—superheterodyne, tuned-radio frequency (TRF), etc. Substantial divergences in detail are found—in bandwidth, tuning mechanisms, etc. Some operational situations justify the continued utilization of receiver types largely displaced in other instances by more common circuits. Thus, direct-detection receivers, superregenerative receivers, etc., are used on occasion and their use is by choice—because they best fit the needs of the job at hand.

(a)

(b)

(*c*)

FIGURE 9-1 Intercept systems. (*a*) A complex reconnaissance center. (*b*) A complex receiver. This system covers 90 megacycles to 10.75 kilomegacycles with 5 rf tuners, shown in the top of the photograph. Two switch assemblies are shown at left center and center. At right center are 2 i-f amplifier units. The power supply is at the lower left and the indicator control at the lower right. (*c*) A "pocket" receiver. This unit covers S-band with a tunable cavity. In addition to the antenna and earphone which are shown, it contains a crystal detector, audio amplifier with gain control, and battery power supply.

This chapter is concerned in particular with microwave intercept receivers. It is not feasible to describe in detail each existing receiver having practical value; there is not the space, and the rapidity with which the receiver picture is changing would outdate such material very quickly. Instead, there first will be repeated from earlier chapters certain material on operational problems, intercept techniques, etc. The justification for the repetition lies in the emphasis here on aspects bearing directly both on the design of intercept receivers and on the selection and employment of particular receivers for some common uses. The objective will be to describe why ECM intercept receivers must and do differ in detail from other microwave receivers of similar basic designs. The material will then relate to this background appropriate interpretations of the common receiver performance parameters—noise-figure, bandwidth, sensitivity, fidelity, etc. The discussions will presume a general knowledge of conventional receivers, circuits, and components; they will emphasize the differences which must necessarily appear in the design and utilization of intercept receivers (Reference 1 and Chapters 23 through 35 of Reference 2). There will be discussions of the several important receiver types currently in use. A listing of intercept receivers is given in Reference 3. And there will be reviews of the principal problems which influence receiver design and of some innovations in circuits and components which point the directions of future trends.

To maintain the general nature of the review, the major attention will be devoted to discussions of types of receivers rather than to descriptions of individual, current receivers.

9.1 Some Factors Affecting Intercept Receiver Design

There are unusual factors, operational and technical, affecting the design of intercept receivers.

(1) *Lack of a priori Information*

Perhaps the most significant over-all distinction between intercept receivers and other types is the fact that the former must operate without *a priori* knowledge of the electronic characteristics or physical location of the signals. Radar receivers, by contrast, have complete knowledge of the signal frequency, pulse repetition frequency (prf), pulsewidth, possibly an estimate of signal arrival time, etc. The initial reconnaissance receiver tasks are to find (intercept) and to identify or recognize the signals in a short time. In ELINT, any concern with message content is often of secondary importance. Further, the majority of intercept receivers must be expected to collect and process a number of unrelated signals simultaneously. The terms countercept

and synchrocept have been suggested to differentiate between the reconnaissance receiver situation and that in which at least the majority of the signal characteristics are known.

(2) *Inability to Use Integration Techniques*

Because of the wide variation in the electrical characteristics of signals, the opportunities for aid to weak-signal detection via signal integration (as in a typical radar operation) are apt to be absent. Built-in integrating devices imply *a priori* information, usually unavailable, or an unusual concentration of interest on a particular type of signal. There is generally a direct signal amplitude versus noise amplitude competition in the detection process such that practical receiver sensitivity (in terms of usable signal-to-noise ratio) is often much less, for example, than values usable in a radar receiver.

(3) *Complexity of Signal Characteristics*

The basic intercept tasks become increasingly difficult with time because of trends in signal character. To avoid detection, or to avoid countermeasure action, modern weapons systems signals are frequently subjected to programmed or even random variations in character (in radio frequency, prf, pulsewidth, etc.) during a transmission interval; the result is to greatly magnify intercept and identification problems. Thus, the value of precise measurement of radio frequency (a prime identifying parameter in ordinary circumstances) must be viewed with some qualification in the era of pulse-by-pulse frequency jump transmission, rapid tuning capability in transmitters, etc. In fact, a receiver having the resolution and stability to measure radio frequency accurately may be at some substantial disadvantage in detecting certain types of signals.

(4) *Divergences in Operational Requirements*

In contrast to electronic countermeasures in general, whose primary value is in tactical missions during hostilities, there is a substantial need for intercept receivers during both "cold-war" and "hot-war" periods. The many differences in physical environment, in signal conditions, in operational use, and in the relative values of different types of data for these divergent applications combine to justify the development of different receivers tailored particularly to the different eras. Even within the cold-war period, there are surprisingly different uses for intercept receivers that dictate the need for a tremendous versatility in receiver design and employment. The demands for intercept information in the cold-war period range from

assessment of an enemy's research and development activities, to monitoring the operational deployment of radar equipments, to determination of intention to attack, to immediate verification of an actual attack. Since the initiation of hostilities may well be in the hands of a potential enemy, a developed capability for all phases of ELINT activity is obviously desirable.

(5) *Divergence in Physical Requirements*

The variations asked for in physical design are tremendous. Thus, receivers are designed for installation in surface ships, submarines, complex ground-based detection centers, motor vehicles, aircraft, reconnaissance satellites, hats, brief cases, etc. The environmental extremes in shock, vibration, temperature, altitude, etc., are fully the equivalent of those imposed on the electronic systems whose signals are to be detected.

(6) *Wide Frequency Ranges to Be Monitored*

In general, the intercept receiver must monitor a total radio-frequency band substantially in excess of the frequency ranges of the individual signals of the electronic systems to be detected within this band. This introduces major technical considerations in wideband circuitry. The d-c-to-light concept was never more applicable.

(7) *Wide Dynamic Ranges Encountered*

The wide variations in received signal level that must be anticipated are enormous. Because of the one-way transmission to the intercept receiver (versus the two-way action that may be involved in the operation of the signal emitting system), signal levels are apt to be high—high sensitivity sometimes is unnecessary (and undesirable because of the possible introduction of lower level interfering signals). But in contrast, the intercept receiver may, in another circumstance, be faced with the task of intercepting a low-power transmission via radiation from minor lobes of a transmitting antenna, and from a great distance—a situation arguing for the maximum sensitivity. An intercept receiver of general utility, then, must be prepared to operate over a very large dynamic range.

(8) *Complex Data Handling Problems*

In situations where a high data rate (occasioned either by a large number of signals to be monitored or by the desire for detailed technical information about selected signals) must be handled, the intercept receiver must frequently operate largely on an automatic or semiautomatic basis. Therefore, the receiver performance versatility

often must be retained in automatic or semiautomatic operation. If the employment of an operator, or operators, is feasible, the consequent reduction in complex automatic circuits may be largely offset by the necessary inclusion of special display and control circuitry leading to the best utilisation of the operator.

(9) *Presence of False Signals*

There must be a continuing and unusual concern with false signals. "False" refers to more than the internally generated spurious responses sometimes encountered in receivers, or to the results of propagation anomalies. It relates to the off-frequency signals generated in high-power transmitting tubes; such signals, while reduced in amplitude far below the normal frequency level, still may represent a very substantial radiation energy. There is always the threat of decoy signals produced by an alert enemy to capture the attention of the intercept systems. There is the threat that certain signal characteristics used to "fingerprint" signals (antenna scan rates, prf, etc.) are being subtly modified by the enemy to lend confusion to the operation. While it is not within the province of the intercept receiver to make fundamental decisions in such matters, it is important that the intercept receiver not introduce further confusion by an inability to handle the received data without further distortion or modification.

Intercept receiver technology is in a continuing state of change. New receivers are developed not only to reflect the recent advances in more versatile circuits and components, but to meet, as well, both new operational uses (in missiles and satellites, for example) and new problems imposed by the ever-changing electrical and environmental characteristics of the signals of interest. Yet, in all cases, there are certain underlying relationships which in some form affect intercept receivers. There are basic operational conditions which affect the choice of receiver circuits. There are limitations, advantages, and compromises in basic receiver techniques which affect the initial design of a particular receiver. There are aspects which affect the selection of a receiver for a particular job, or which determine the optimum employment in field use of the selected receivers. These are the factors considered in this chapter.

9.2 Two Important Operational Requirements: Intercept Probability and Signal Selection

The principal job of the intercept receiver is to provide information on the existence and nature of various signals—usually in the minimum pos-

sible time. The detailed questions asked of an intercept system might be one or more of the following: Are there any signals present? What are the electrical characteristics of and directional bearing to those signals present? Is there a signal present having certain prescribed characteristics (perhaps in frequency, pulsewidth, prf, etc.)? Is there a signal present which is tracking the location of the receiver? Is there a new signal present (new in the sense that the transmission has just been added to the general signal environment)? Is there an unusual signal present (unusual in the sense that it possesses characteristics not found in the current catalog of signals)? Is there a signal present that evidences certain characteristics of motion (perhaps identifying a missile or aircraft transmission)? Are there c-w signals, FM signals, single sideband (SSB) signals? Is there evidence of simultaneously or sequentially pulsed transmissions on adjacent frequencies? In adjacent bands? In widely separated bands? Is there a particular single signal in existence the mere presence of which conveys important information? Is there a change in the general pattern set up by many signals (increase or decrease in density, general geographical disposition, etc.) which carries immediate tactical implications, perhaps as to an imminent missile firing, a realignment of ground forces, a relocation of a base, etc.? The list is almost endless. It is apparent that no one intercept system will answer all such questions optimally.

Regardless of the specific questions appropriate to the problem at hand, there are certain general aspects of the intercept that always must be considered: (1) the nature and amount of information to be developed for each problem signal (knowledge of mere existence versus determination of detailed characteristics), (2) the forms in which the output data are to be provided (lamp indication to an operator or camera; panoramic presentation; an electrical output suitable for computer processing, for recording, for retransmission, or for control of other circuits, etc.), (3) the time available for data interception and processing (very short in the case of initial attack warning versus relatively long in the case of certain cold-war reconnaissance and monitoring processes), (4) the environment in which the receiver must operate (this encompasses both the physical environment and the signal environment).

Many of the practical implications of the above items in receiver design arise in the consideration of two factors closely identified with intercept receivers. The degree of *intercept probability* defines the basic ability of the receiver to provide, within an acceptable time interval, reception of a transmission; the term considers the typical elapsed time between the initial existence of a detectable signal and the initial reception. *Signal selection* relates to the further ability of a receiver to separate a signal for additional

treatment, and possibly to identify an intercepted signal as being of a certain type or class. It refers to abilities to measure signal characteristics, and/or to select signals on the basis of electrical characteristics or operational behavior as being ur. 'e and, therefore, worthy of particular interest. Note that the term intercept probability is frequently broadened in meaning to imply the ability to identify signal characteristics as well as to determine mere existence. This is particularly logical in discussing reconnaissance receivers. However, a separation of functions (into intercept probability and signal selection) is useful in considering the technical details of the receivers. These topics have received special treatment in Chapters 4, 5, and 6; the concern here is with their influence on receiver design.

9.2.1 Intercept Probability

A basic requirement for the principal job of the intercept receiver is *high intercept probability*, and no other single design objective has exerted such influence on the development of intercept receivers.
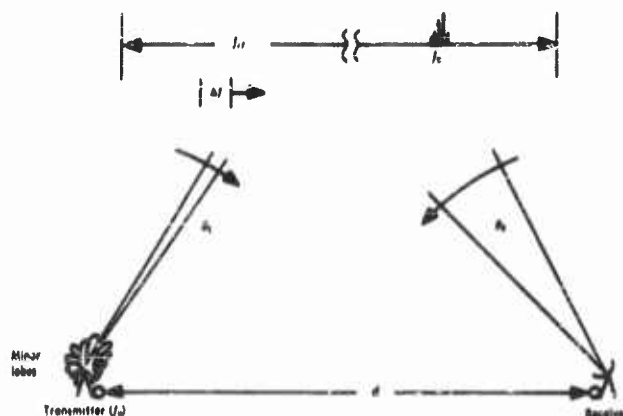
FIGURE 9-2 A typical radar intercept problem. Scanning antennas are involved at both transmitter and receiver sites; the receiver scans an incremental acceptance bandwidth Δf across a total monitored frequency range of f_{rf}. Under direction and frequency (f_0) coincidence conditions, a limited number of radar pulses would be intercepted.

A basic radar intercept problem is outlined in Figure 9-2. Therein is depicted a tunable intercept receiver operating with a programmed change (scan) in frequency (an incremental acceptance bandwidth Δf is tuned periodically through f_{rf}) and utilizing a rotating directional receiving antenna. The radar is presumed transmitting on an unknown frequency (f_0) within the tuning range of the receiver (f_{rf}) in a varying direction as con-

trolled by its rotating antenna. For intercept, the frequency scan process must have tuned the receiver to the proper frequency at a time during which a signal of detectable amplitude is delivered to the receiver location. This is most likely to occur when the receiving antenna is aimed at the radar and when, simultaneously, the radar antenna is scanning through the bearing to the receiver. These conditions must persist for a time long enough for the receiver to establish the identity of the signal as such. For tentative identification of conventional radar transmissions, this may simply mean that the receiver must not tune through the radar frequency in the interval between pulses. If knowledge of signal prf is required, the receiver must remain receptive to the signal long enough to intercept at least two consecutive pulses. If programmed or random variations of emitter characteristics are to be anticipated, the requirement encompasses correspondingly longer times.

Note that there are several scanning processes involved. The probability of the necessary time-frequency-direction coincidence is indeed small. As the scanning processes are reduced in number, the time probability situation improves rapidly; however, it is inevitable that as the number of scanning processes is reduced, the sensitivity of the system is also reduced, unless corresponding complexities are incorporated (multiple antennas, channelized front-ends, etc.). In comparing the relative merits of various choices and combinations of scanning processes, it is important to consider not only the sensitivity and degree of signal selection attainable, but the possible restrictions that may be imposed on the quality of signal information (pulse shape, modulations, etc.) and on the receiver data rate (number of signal characteristics than can be measured, number of signals that can be simultaneously monitored, etc.). Fortunately, it is frequently possible, at least for initial intercept, to use an omnidirectional receiving antenna ($\theta_a = 360°$). Further, it may be possible to use a non-scanning receiver wide-open in frequency such that it continuously monitors the entire assigned r-f band ($\Delta f = f_{rf}$). Only one scanning process then remains (that of the radar antenna) and a signal of sufficient amplitude would be detected as soon as the radar antenna scans through the receiver bearing. Alternatively, a narrow-band (narrow Δf) frequency scanning receiver might be used; this presumes a higher sensitivity such as to assure the capability for detecting the radar transmission through minor-lobe radiation. Detection, then, would not require the radar beam to be directed at the receiver. Again, a single scanning process remains (that of the receiver in frequency) and the radar could be detected as soon as the tuning program adjusted the receiver to the radar frequency.

The situation in which the radar signal is detected through minor-lobe

radiations or through energy scattered from site obstacles is common. The use of this intercept process is implied as optional; its use frequently is mandatory if there is some strong likelihood that the radar transmission would not periodically be beamed at the receiver (as is true with some weapons systems, or at some receiver locations). Fortunately, field measurement data indicate substantial radiation from minor antenna lobes and from
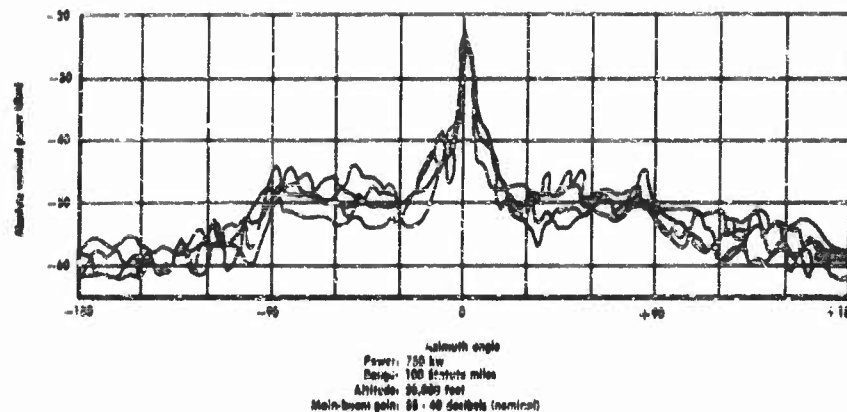
FIGURE 9-3   Minor-lobe radiation from 4 S-band radar sites. (Source: References 36 and 37.)

site reflections in many typical cases as shown in Figure 9-3. The same is generally true as regards radiation toward the zenith, a situation of importance in high-altitude reconnaissance. Again, measurements show average zenith radiation from a typical ground-based radar antenna to be, commonly, only a few decibels below that which would be radiated from an isotropic antenna. Some interesting results of such measurements are contained in Reference 4.

The variations to the general intercept problem are many. The radar transmission may be of very short duration, the frequency might not be constant from pulse to pulse, etc. Much of the design philosophy of intercept receivers is based on the need to adjust conditions (scan processes, bandwidths, etc.) to handle optimally the many variations.

Simply stated, a high intercept probability exists only when there is a high probability of intercepting a particular signal, or group or class of signals, in an operationally acceptable short time. Thus, a simple, broadband crystal-video receiver of low sensitivity might provide the optimum intercept probability in a bomber defense problem where practically instant warning of a strong, local signal is all that is desired. Yet that same receiver

would have no practical intercept probability in general reconnaissance for distant signals too weak to be detected by the crystal-video device. For the latter task, a slowly tuned superheterodyne having a narrow acceptance band (and a consequently higher sensitivity) might provide at least a possibility of success in detecting the weak signal. Yet the superheterodyne might be relatively useless (because of the slow scan in frequency) in the apparently more simple warning task in bomber defense where time is of the essence.

High intercept probability in an intercept system implies, first, a sufficient sensitivity in the receiver to permit detection of the signals of interest. Beyond this, it implies particular abilities to insure detection. It guarantees certain r-f bandwidth and tuning characteristics such that there is a high probability that the receiver will be receptive to the frequency of transmission early in the transmitting interval. It suggests that beamwidth and scan characteristics of the associated antenna will favor reception from the direction of the signal during the transmitting interval, and it indicates that there are characteristics in the receiver aiding the recognition of a signal, as such, in the presence of noise and other signals.

9.2.2 Signal Selection

Historically, an intensive search for special receiver techniques some years ago produced several important advances in receiver design contributing to higher intercept probability. This general success, in turn, has emphasized an accompanying problem—that of segregating, analyzing, and recording the wealth of data developed in a high-signal density environment by the "improved" receivers. Thus, along with the more obvious general receiver design requirements of (1) a satisfactory intercept probability for the signals of interest, and (2) a design consistent with the nature of the physical environment in which the receiver will be used; there is now a general recognition of the possible need for (3) a special ability to operate in a prescribed high-density signal environment, and (4) a design consistent with particular requirements for data read-out, analysis, recording, etc.

There are several design approaches to items 3 and 4. If reconnaissance for strategic intelligence purposes is the primary objective for the intercept receivers, an attempt is sometimes made to develop a receiving equipment of such versatility that it can handle all signals in a complex environment; the intercepted data are processed and cataloged immediately in great detail by the field equipment. Alternatively, the fidelity of detected signal reproduction of the receiver is made such that, in conjunction with a wideband recorder, a record of the full signal environment can be made with an accuracy and detail such as to permit subsequent analysis under more con-

venient analysis conditions. A major value of this second approach lies in the ability to re-examine the data in a near-original form in order to answer secondary questions that might later be posed. An important problem may remain if the recording must be restricted to envelope characteristics; signal identifying features contained in small unintentional variations in radio frequency or phase might be lost, as well as intentional variations that are fundamental to the operation of the intercepted system (FM, pulse compression, etc.).

These are standard approaches of obvious validity; they need no further discussion. They are safe in that the chances of missing an important bit of data are minimized since, theoretically, all signals are examined and treated. However, these approaches, of necessity, require complex field equipment. These schemes are sometimes combined in complex systems; the usual intention then is to record only those unusual signals which require further analysis.

A useful alternative approach can be exploited if a less general reconnaissance task is at hand, or if the receiver is to be used as a warning device, or as a signal selector in a jamming operation. Some signal presegregation can be employed, commonly by an electrical presorting within the receiver itself, such that the output contains only those signals which meet certain prescribed characteristics—in frequency, pulsewidth, prf, polarization, modulation, antenna scan, amplitude, etc. (Only two or three of these parameters are ordinarily examined in a single equipment.) The details of these few selected signals are then read out or recorded.

In a sense any scanning receiver which has an incremental bandwidth less than the total r-f spectrum assigned for monitoring necessarily does "signal sorting" in that its attention at any time is restricted to less than the total picture; it does not receive all signals all of the time. However, this fact is the basis of a major intercept probability *problem* in receivers employing frequency scanning, and such frequency selectivity may not, in those circumstances, be a useful presorting technique since the resulting reduction of data is on a substantially unplanned basis. In particular, intercept and analysis of the significant parameters of a modern radar utilizing rapid variations in transmitted frequency would be far from optimum.

The objective of signal sorting is to reduce in a logical, selective manner the amount of data that must be subsequently treated by an operator or by associated analysis, recording, or control circuitry. The success is dependent on how well a signal, or class of signals can be uniquely identified by a few sorting parameters; the job obviously becomes more complex in high-signal density areas. Fortunately, certain types of signals can be selected quite simply via the use of two or three sorting parameters such that remarkably
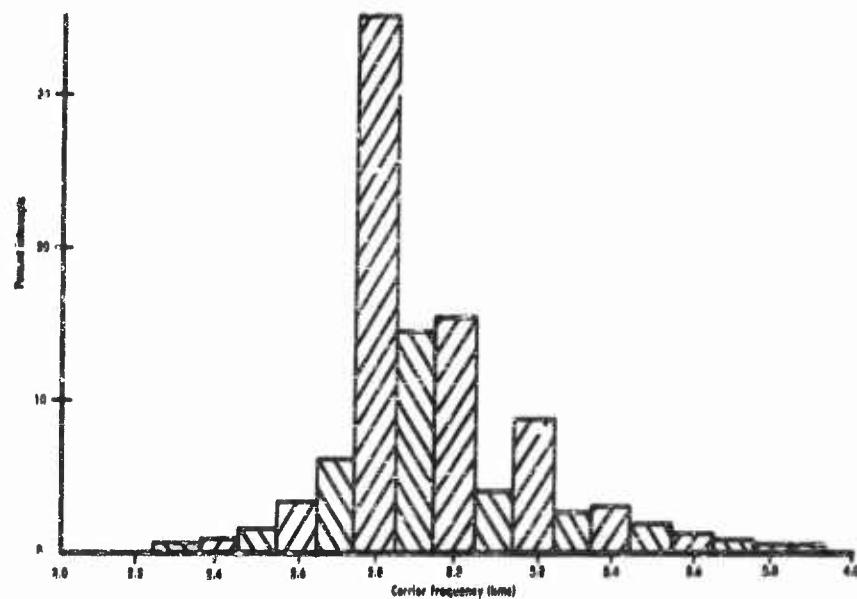
FIGURE 9-4   Representative S-band frequency distribution.
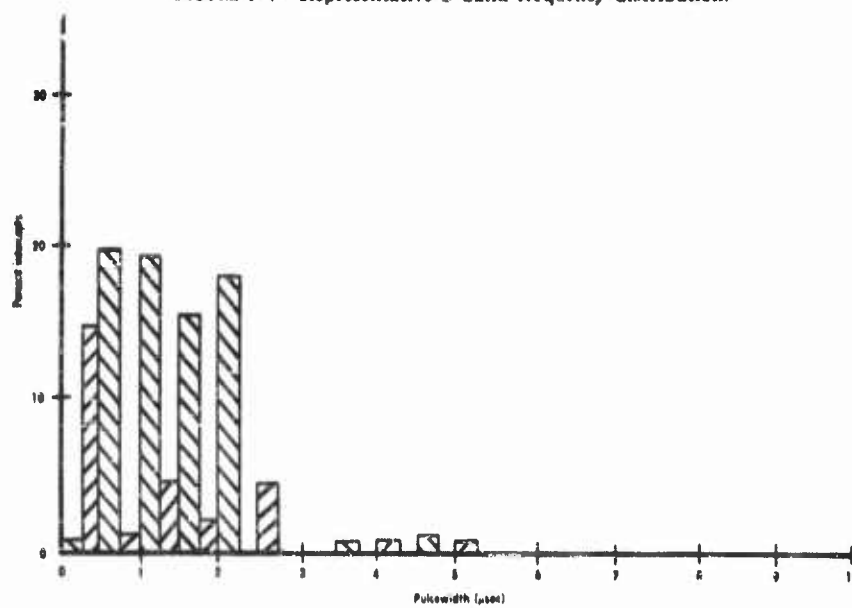
FIGURE 9-5   Signal distribution as a function of pulsewidth.
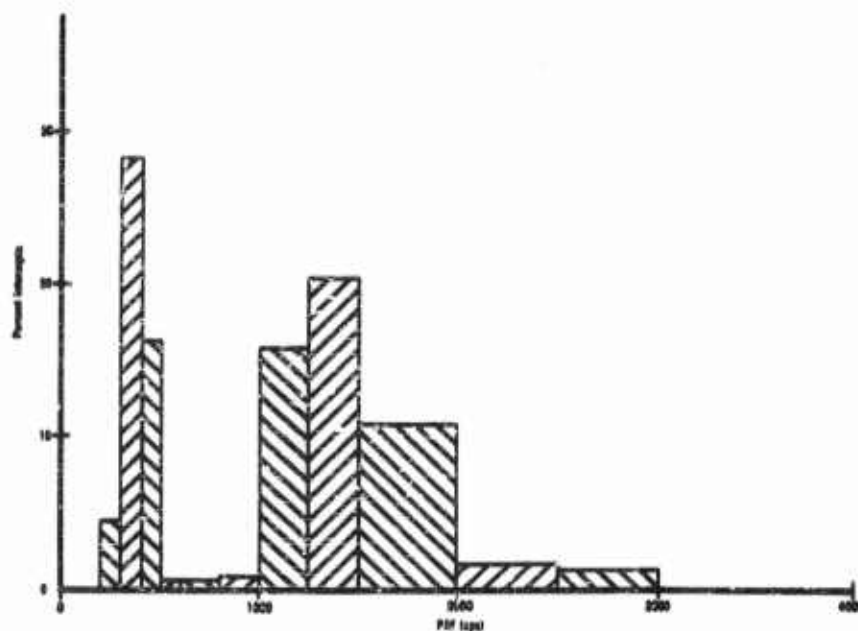
FIGURE 9-6 Signal distribution as a function of prf.

simple receivers (but of restricted general utility) can be devised for special purposes.

There are restraints imposed on the practical utility of the various schemes for presorting, presegregation, automatic signal recognition, signal identification, etc.

There is first the concern with whether or not a signal can be uniquely identified through measurements of a limited number of characteristics. For some signals there is no basic problem; but other important signals may differ but little from the many signal types which compose the total environment. Signals tend to group together, particularly in frequency, as is evident in the plot (Figure 9-4) of the frequency disposition of 90-odd S-band signals intercepted in a short flight along the California coast. Fortunately, the grouping is not nearly so pronounced in certain other characteristics —illustrated for pulsewidth and prf of the same S-band signals in Figures 9-5 and 9-6 respectively. Figure 9-7 is illustrative of the general sorting problem and shows the cataloging of S-band signals into a number of bins that would be brought about by sorting through measurements of several common signal parameters. (It should be noted that two parameters
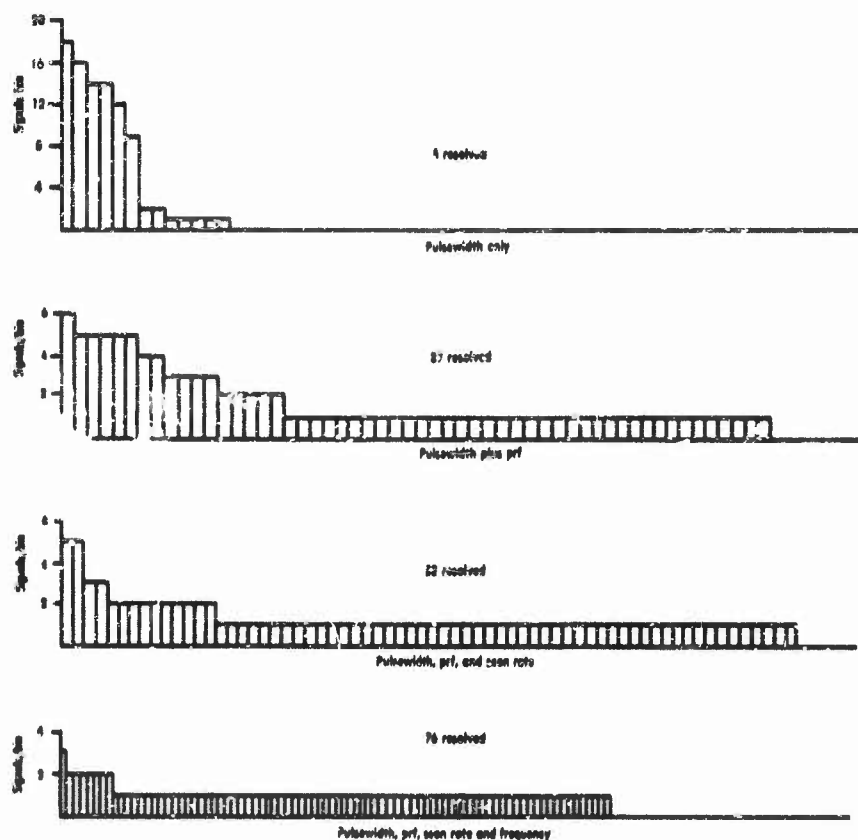
FIGURE 9-7 Signal sorting brought about by sequential sorting processes (93-signal S-band sample). Pulsewidth accuracy: 10%; prf accuracy: 6%; scan-rate accuracy: 12%; frequency accuracy: 50 megacycles. Individual parameter resolution capabilities: prf, 3 signals; scan rate, 5 signals; frequency, 7 signals.

may not be really independent; pulsewidth and prf are often not.) The objective of the sorting is, of course, to have as few signals per bin as possible. Not evident in Figure 9-7 is an important additional "sorting in time" brought about automatically by the scanning of the directional antennas associated with the signals and by the travel of the receiver (the sample was gathered by a receiver in an aircraft).

A second aspect of the sorting process, once the parameters to be sorted have been chosen, is the order in which they can optimally be handled in an intercept receiver. Certain sequential arrangements are usually suggested by the time required to measure each parameter, but important alternatives involve simultaneous measurement of certain combinations with perhaps a

coincidence of values of these parameters required before further sorting is attempted. For example, frequency and direction can be obtained on a pulse-by-pulse basis and a subsequent correlation of both of these parameters demanded by the equipment before prf (or interpulse interval) is measured. The simplification in prf measuring circuitry resulting from this presegregation of interlaced and unrelated signals can be considerable.

There is a third aspect, one over which the equipment designer has little control—the quality and quantity of signal data available to the receiver. This affects system operation regardless of inclination toward treating all data or only presorted data. Propagation is a major factor in distorting the data available to the receiver for measurement. Figure 9-8 shows a 1-micro-

Time →
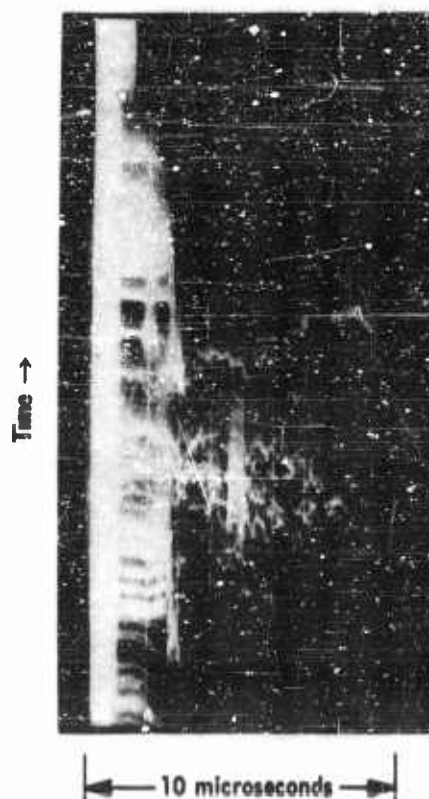
|◄—— 10 microseconds ——►|

FIGURE 9-8 Pulse "trains" produced by terrain and site reflections. Successive "A-scope" traces are shown with a 10-microsecond time base. The overlapping direct pulses produce the solid paint in the initial microsecond. The scanning of the S-band radar beam produces the variable reflections.

second pulse as received at a ground site in combination with reflected energy from terrain features. If the time delays resulting from multiple-path propagation are greater than a pulsewidth, the resulting signal is received with one or more distinct echoes. The energy level of these echoes is much lower than the direct path signal if the radar antenna is directed toward the receiver. However, when the receiver sensitivity is sufficient to receive minor lobes of the radar via direct path propagation, the corresponding reflected echoes of the main lobe can have equal or greater energy. Thus, extraneous pulses can be received whose width will, in general, be different from that of the transmitter. Without adequate precautions in the equipment, these can cause both false measurements and incorrect implications about the number of emitters present. (Site and terrain reflections account in large part for the substantial minor lobes associated with the S-band radar antenna pattern plotted in Figure 9-3.) A pulse is shown in Figure 9-9 following trans-
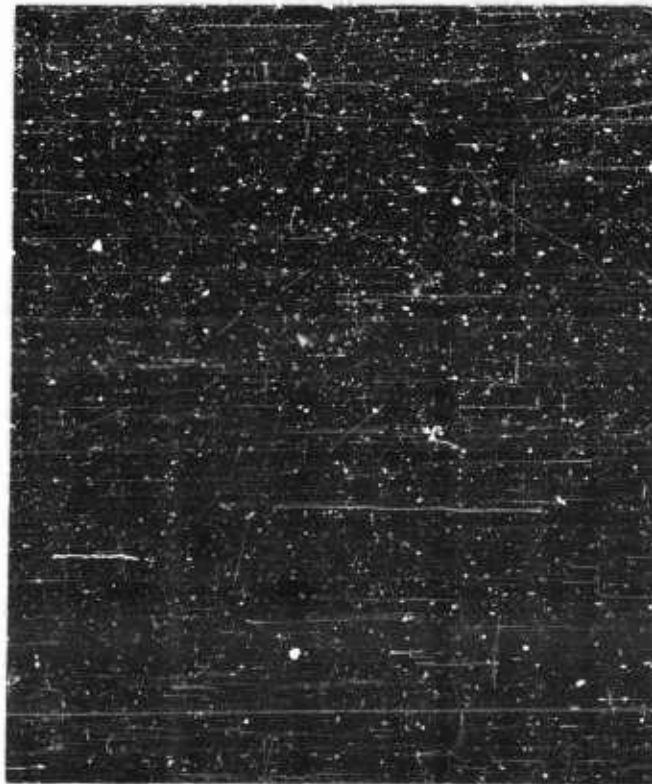
FIGURE 9-9   Effect on pulse shape of multipath propagation between fixed points. Successive 2-microsecond S-band pulses are shown in three groups displaced in time by a few seconds.
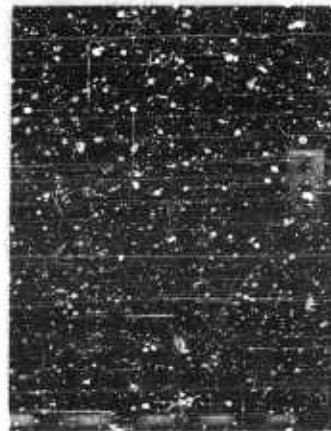
Figure 9-10 Bearing variation in beyond-horizon S-band scatter signal as a function of time. Four 5-second samples are shown. A 0.5° beam scans a 3.8° sector each 0.1 second. The true bearing is aligned with the centerline of the trace.

mission through trees; multipath distortion effects on shape are evident. The apparent time variations in the bearing to a transmitter involved in an S-band transhorizon transmission via tropospheric scatter are shown in Figure 9-10; the successive traces are only 0.1 second apart. Signal density often appears as a major problem, the quantities of signals with which the circuitry may have to contend in some instances are staggering (References 5 and 6).

Presuming there is some justification for attempting a presorting of signals, a third question then arises. What is the necessary measurement accuracy? An important control on refinement of receiver measuring abilities is set by the value that can be attached to the readings. Qualifications are introduced by the factors outlined earlier—the spurious signals, etc. In situations requiring emitter parameter details as a matter of technical intelligence information *per se*, absolute accuracy of measurement may be significant even if only one signal is received at a time. A high degree of measurement accuracy might also be required in correlating intercepts made from different geographical locations or recorded at different times. In other situations, where a capability for presorting signals is the prime consideration in establishing accuracy requirements, some latitude may be permissible. The discussion to follow is concerned with the implications in presorting signals, and other circumstances in which it is legitimate to question the need for high measurement accuracy and its attendant cost in equipment complexity. If all pulse signal repetition rates were fixed—say 300 pulses per second—there would be no need to measure the repetition rate. If there were only two rates, perhaps 300 pulses per second and 2000 pulses per second, the measurement accuracy could be very coarse indeed, yet be completely useful in identifying the proper one of the two categories. The accuracy need be no greater than that required to establish meaningful differences. However, these differences are sometimes small, for example, when it is desired to identify a particular transmitter unit of a class by small differences in signal parameters—to fingerprint a radar by noting that its

pulse repetition rate is 299 pulses per second rather than the normal 300. This process legitimately calls for a high relative measurement accuracy; the resulting information can be of substantial value when there is some certainty that the minor parameter perturbations do not vary with time (by chance, or by the design of some uncooperative victims).

Finally, can sorting of the precision required be accomplished technically with reasonable circuitry and in a reasonable time? There are genuine problems in devising automatic circuitry which measures pulsewidths over very wide dynamic ranges; which relates properly the pulses in interlaced pulse trains; which measures signal polarization in a definite way; which determines antenna scan rates or scan patterns; which defines modulation types, etc. Fortunately, the precision required in the measurement of any one quantity drops as the number of signal parameters taken in combination in the identification is increased. The measurement accuracies tabulated in Figure 9-7 are typical and are not unreasonable in simple equipments.

This section has emphasized some problems faced in certain aspects of intercept receiver development and utilization. The purpose is to account for and justify some unusual practices in intercept receiver circuitry. This recounting of so many problems may carry implications as to the value of the intercept operation. There is none intended; there are ample demonstrations of the utility of intercept receivers and of the practicability of deriving complex information of great value through their use.

9.5 Intercept Systems

Section 9.2 introduced a number of typical questions asked of an intercept receiver, such as: Is there any signal present? What are the electrical characteristics of the signals? In reality, these are translations of much broader questions that are sometimes important far from the field of electronic reconnaissance or warning. Some of these might be: What is the disposition of an enemy's land or naval forces? What is the electronic order of battle? What is his ability to defend or attack? What is the state of his technology? What are his intentions about surprise attack? Has an attack been launched? Is there a weapon system trained on an aircraft or submarine?

The answers to such questions may be provided at least partially by signal intercept alone; ordinarily, however, an intercept *system* will be required for the total job—to provide all possible primary data, to translate the data into meaningful answers, and to transfer the answers to the intended user in an acceptable time.

The basic intercept job may be done with simple equipment—antenna, receiver, and indicator lamp. Sometimes a much more complex assemblage of subsystems is required. In any event, the intercept receiver alone is the concern of this chapter. In practice, the influences of the characteristics of the antenna and of the associated processing and display circuitry cannot be

Ignored in intercept receiver design and utilization. The importance of these matters justifies treatment in a full chapter; accordingly, antennas and their influences are discussed in Chapter 29, direction finding is covered in Chapter 10, and analysis and display problems are treated in Chapter 11. . one aspects of a system are subject to some control in equipment design and utilization. Propagational effects here offer some contrast but may well be of equal importance in the successful utilization of a system—particularly in those circumstances involving diffraction and scatter effects. These matters are covered in Chapter 31.

9.3.1 Receivers as a System Component

Beyond the simple system as visualized above (antenna, receiver, and indicator), the complexity of certain intercept problems justifies the utilization of substantially more complex systems. Correspondingly, more complex receivers appear in such systems. Ordinarily, the r-f spectrum to be monitored will exceed the tuning range of a single receiver. Thus, an intercept receiver might have several tuning heads (covering different r-f bands); or if simultaneous monitoring of the several bands were of paramount importance, several receivers might be devoted to the total system problem. Each would be associated with its own antenna, and there might be substantial differences in individual receiver design as required to develop acceptable characteristics in the different r-f ranges. It is likely that some interrelation of the output data would be employed via a composite read-out unit.

Even if only a single r-f band were to be monitored several receivers might be employed in combination in a system. It is not uncommon to utilize a guard (alerting, warning) receiver featuring special intercept capabilities first to detect quickly the presence of a signal, and then to aid the utilization of a second, analysis (precision) receiver for detailed signal inspection and read-out. A third, d-f receiver is sometimes employed with an associated directional antenna. Several receivers might be incorporated into a system with the division of utilization to be based on automatic, semiautomatic, and manual operation. This usually brings about a corresponding segregation of data such that the numerous routine signals are handled automatically, the fewer, unusual signals receiving a more detailed treatment in accordance with the abilities (and limitations) of the semiautomatic and manual processes.

Systems problems in the full sense arise in such circumstances. There are the usual technical problems associated with integration of equipments, compatibility with the environment, logistics, interfaces with other systems (a data transmission system, for example), etc. But the system concept often must be employed simply because an intercept receiver, by itself, is incapable of providing the desired answers to questions of strategic or tactical importance.

As an indication of the potential (but necessary) complexity of intercept systems, one system proposed for long-range, ground-to-ground detection involved, per self-sufficient site, 24 intercept receivers, a total equipment weight of many tons, 12 men in the actual operating crew alone, and an initial cost (installed) of $15,000,000. An airborne reconnaissance equipment (covering several bands and possessing some remarkable abilities) weighs 4500 pounds and costs in the order of $2,000,000 per unit. Fortunately, the simple system shown in Figure 9-1(c) is also of substantial operational value.

9.3.2 System Evaluation in Terms of Signal Intercept

It is often necessary that receiver operational characteristics be evaluated with respect to the total system objectives. System intercept probability is influenced by several receiver and antenna characteristics—noise-figure, gain, predetection (acceptance bandwidth), and postdetection bandwidth, tuning range, tuning program (frequency scan characteristic), antenna gain, beamwidth, and scan characteristics, etc. It is difficult to relate these factors to give some quantitative measure of the intercept probability of a system. But there is a very general relationship leading to a figure-of-merit based only on the composite abilities of an intercept system to monitor simultaneously both frequency and geographic area (more properly, volume-of-space). This concept presumes that the frequency and location of a signal are unknown and that there is an equal profit to success in intercepting a signal to be brought about by doubling the frequency band monitored per unit of time or by doubling the volume of space (or surface area) monitored in that time interval.

The common denominator to improvement in both aspects is receiver noise-figure. This is because an improvement in receiver noise-figure is the only step of unquestionable value in providing for a basic improvement in sensitivity without some penalty to another performance parameter. Sensitivity, as such, can be bought without a change in noise-figure by a constriction of the receiver's effective noise bandwidth. The attendant increase in maximum range of reception for a signal of given characteristics results in a greater space volume monitored per unit of time; but the accompanying reduction in acceptance bandwidth of the receiver decreases the frequency range monitored per unit of time so that there is no net improvement in the figure-of-merit. In similar vein, system sensitivity can be purchased with higher antenna gains. The accompanying range increase would appear to improve the space coverage of the system but the beamwidth reduction associated with the increased gain so obtained balances out any improvement in actual space coverage per unit of time. In contrast, a noise-figure reduction can increase sensitivity—and range—without any necessary reduction in bandwidth or beamwidth.

Considerations of *time* are inevitable in a discussion of system intercept probability. The figure-of-merit referred to above measures system performance on the basis of frequency and space coverage per unit of time. A minimum value for this unit-of-time is set by the interval over which a signal must be observed in order to establish its existence or in order to measure certain characteristics. It varies with the task at hand, a few milliseconds often sufficing for pulse signal intercept in the microwave bands. If the time available for intercept is longer, i.e., if the transmissions exist for several "units" of time, a system of lower figure-of-merit might be able to accumulate more useful data over the total time interval. This situation is illustrated in Figure 9-11 where the relative intercept abilities of six hypothetical receiver-antenna systems are contrasted. The systems are assumed to be employed in a ground-to-ground monitoring operation, whence an area-of-space (rather than a volume-of-space) is the configuration of interest; all receiving antennas are assumed for the sake of comparison to have a fixed vertical beamwidth of 20°. To simplify the intercept probability problem somewhat, it is assumed that side- and back-lobe radiation from the emitters of interest must be received; the beam power of all emitters is assumed to be 1 megawatt or more, with an average minor-lobe antenna gain of 0.2 (7 db below isotropic) giving a minimum effective radiated power of 200 kw. It is further assumed that an azimuthal sector of 150° and a frequency range of 2 kmc at S-band must be monitored by each of the receivers compared. The unit-time interval (required to recognize a signal) has been chosen as 1/50 second, and it is assumed that this length of time is devoted to each incremental azimuthal beamwidth and to each incremental acceptance bandwidth in the scanning process. The so-called system figure-of-merit, as provided by an arbitrary multiplication of frequency coverage (in megacycles) and area coverage (in square miles) is plotted along the abscissa. Receiver A is wide-open in frequency and employs an omnidirectional antenna. The figure of merit (performance number) of this combination is given by the intercept of curve *A* with the abscissa, this marking the performance for the initial unit-time interval. It is unchanging with time (except for the practical benefits of the continuing monitoring) since there are no scanning processes incorporated in the receiver or antenna operations. In effect, the simple receiver can determine only the existence of a signal and, perhaps, pulsewidth, prf, and transmitting antenna scan rate. The second combination employs a directional antenna. No performance change results for the initial unit of time; despite the greater range, the *initial area* coverage is unchanged. But the over-all performance improves with antenna scanning (because of the area increase) as indicated in Figure 9-11 and reaches a much higher eventual performance number by the end of a complete antenna rotation (in 30 units of time). Some d-f capability is added. Thereafter the
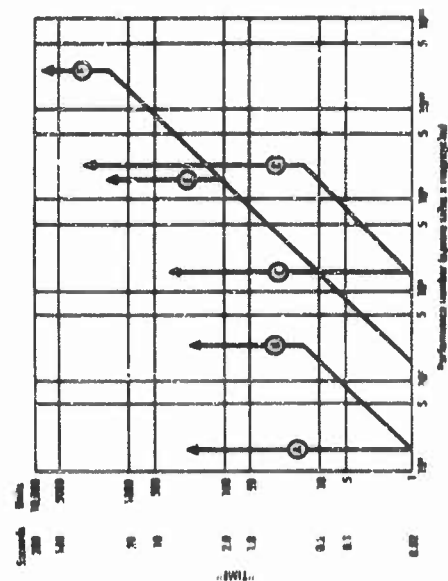
Figure 9-11　Comparative performance chart. Total rf range monitored is 2000 megacycles at X-band. Total azimuth monitored is 150°. A ½-second interval is spent on each incremental bandwidth and on each incremental acceptance bandwidth in the scanning process; i.e., 15 time units (0.3 seconds) are required to scan a 10° antenna through 150° for each incremental frequency bandwidth. All antennas have 20° elevation pattern beamwidth. The weakest signal of interest is minor-lobe radiation (7 db below isotropic) from a 1-megawatt radar.

Curve	Receiver Type	Sensitivity (dbm)	Acceptance Bandwidth (Mc)	Antenna Beamwidth (degrees)	Initial Performance Number	Final Performance Number	Total Scan Time (sec)	(units)
A	Crystal video	-50	2000	150	1.64×10^8	1.64×10^8	0.02	1
B	Crystal video with directional antenna	-50	2000	10	1.64×10^8	2.45×10^8	0.3	15
C	Preamplified crystal video	-70	2000	150	1.64×10^8	1.64×10^8	0.02	1
D	Preamplified crystal video with directional antenna	-70	2000	20	1.64×10^8	2.45×10^9	0.3	15
E	Tunable with omni-directional antenna	-80	20	150	1.64×10^9	1.64×10^9	2.0	100
F	Tunable with directional antenna	-80	20	10	1.64×10^9	2.45×10^{10}	30.6	1530

operation is, to a degree, redundant. In a similar manner, the addition of a broadband TWT preamplifier provides an over-all benefit, as indicated by curve C and curve D (which shows the effect of the addition of a directional antenna). The 5th and 6th combinations (curves E and F) involve scanning superheterodyne receivers with relatively narrow (20 megacycles) acceptance bandwidths. It is interesting to note that the initial performance number for these two receivers is less than that of the TWT plus wide-open crystal-video receiver despite the higher sensitivity of the superheterodyne. Combination C might be the preferred choice as a warning receiver for a very short duration signal. But if the signal transmission were of longer duration—preferably greater than the scanning times of the scanning systems—the total abilities of the scanning superheterodyne systems to develop information eventually would exceed that of the simpler receiver (the relative limits are tabulated on Figure 9-11); the scanning systems then would provide the better results, i.e., there would be a probability of intercepting transmitters when located at greater distances. Further, a frequency measuring ability has been added.

The practical value of the figure-of-merit concept is limited; but it does support the idea that there is a matter of compromise in system design involving a *trade* of abilities. Sometimes the profits or penalties are obscure. For example, high altitude reconnaissance might be conducted with a scanning receiver located in a moving vehicle. It might be necessary to use a downward-directed directional antenna, perhaps to develop sufficient sensitivity to detect weak, minor-lobe zenith radiation. The system, then, is constricted both in instantaneous frequency coverage and in instantaneous space coverage; both frequency and space scanning processes are involved (the latter is brought about by the motion of the vehicle). An investigation of the figure-of-merit of the system in terms of a general intercept probability variation with respect to altitude and speed of the vehicle shows no change in figure-of-merit over surprisingly large ranges when the system is optimally designed for each combination.

Of course, there are many reasons arguing for the selection of a particular receiver acceptance bandwidth, system antenna beamwidth, etc., other than the consideration of an idealized general intercept probability. It might be that all expedients in bandwidth and beam narrowing might have to be exploited simply to generate enough sensitivity to provide some small possibility of intercept (a not uncommon situation at frequencies above X-band). Or it might be that a narrow acceptance band is justified by stringent frequency resolution requirements, etc. This discussion is intended to display contrasts in the abilities of some common receiver-antenna combinations, and to point up the sometimes unsuspected merits of some of these combina-

tions—the broadband crystal-video receiver plus low-noise rf preamplification, for example. The major influence of the antenna is also evident, and although such matters are not within the purview of this chapter, the relative merits of omnidirectional and directional antennas for signal-search purposes, and the effects of beam width and rotation rate of directional antennas on intercept probability become apparent.

Choices of design parameters of receivers must be made in the knowledge that there may be some attendant penalty to system figure-of-merit and that there is an optimum choice—optimum in the sense that what is sacrificed in space coverage or in frequency coverage or in monitoring time can best be afforded. The choices are different for different operational situations and the many useful combinations of receiver parameters result in the many existing forms of receivers found in intercept systems.

9.4 Sensitivity Standards and Signal Strengths

Several standards for quoting microwave intercept receiver sensitivities have come into common usage. All define sensitivity in terms of a minimum acceptable signal power, usually specified in decibels below a milliwatt (−dbm) delivered to the receiver input terminals. Frequently in the practical case there is a subjective decision involved as to whether a signal pulse could (or would) be detected in the presence of the inevitable noise. There are influences imposed by the type of signal, the type of decision (visual display plus operator, automatic electrical detection, etc.), the time that can be devoted to the decision, etc. It is obviously important in comparing receivers to know that identical standards for quoting sensitivities are being employed, and that these standards recognize the practical needs of the intercept job. Section 9.4.1 will discuss three standards which have the virtue of reasonably definable relationship between minimum signal power and the receiver noise power with which a signal must compete. These are (1) equal signal-to-noise standard, (2) tangential signal sensitivity, and (3) triggering sensitivity.

9.4.1 Sensitivity Standards

A problem in intercept receivers stems from the lack (which must be presumed in most cases) of *a priori* information concerning the character of the signals to be intercepted. The receiver usually must accommodate simultaneously a variety of signal types—perhaps the design cannot be made optimum for any one. As a consequence, a signal-to-noise ratio may well be required that is higher than that acceptable in radar receivers where integration is possible. It is a rare case when a microwave signal whose peak power merely equals average noise power can be detected. However, sensitivities are sometimes quoted on that "equal signal-to-noise" basis.

It is more likely that the intercept receiver will require peak signals many decibels above the average noise power. Thus, it has become common, alternatively, to rate receivers as to "tangential sensitivity." This is a term referring to a signal-to-noise situation such that the presence of a signal pulse will raise the noise level by an amount equal to the average undisturbed noise level as viewed on an A-scope presentation. This also presumes that such an event can be noted in practical circumstances. The situation is as

indicated in Figure 9-12. An increase in signal power of about six decibels above the equal signal-to-noise condition is required to establish the tangential-signal condition. In other words, a receiver with a tangential signal sensitivity of −84 dbm is actually 6 decibels more sensitive than one with a quoted −84 dbm equal signal-to-noise power sensitivity.

FIGURE 9-12 A-scope presentation of "tangential signal."

With the increasing interest in automatic detection and operation, a triggering sensitivity becomes particularly meaningful. This term recognizes the incorporation of decision circuitry (in place of a human operator) in the system to make the basic decision as to the presence or absence of a signal in a given observation time interval. Much subtlety has gone into the design of "electrical detectors" to enhance their abilities in signal recognition. To the extent something is known of the signal (its bandwidth, modulation form, etc.) certain advantages can be gained. But in the general case of deciding that a particular observed pulse represents a signal and not a random noise incident, little is ordinarily done except to base the decision on amplitude. If a pulse exceeds a certain amplitude with respect to the average noise level, there is a certain probability that it represents a signal; the higher the amplitude, the greater will be the confidence in the decision. If a low false-alarm rate must be established, the triggering threshold must be high and, typically, might require signals 15 decibels above the average noise level. Thus, a receiver designed for electrical detection might have a triggering sensitivity of −75 dbm. In a sense, this would represent an engineering equivalent of a receiver with a −90 dbm sensitivity based on equal signal-to-noise conditions, or of a receiver of −84 dbm sensitivity based on tangential sensitivity standards. Note that −75 dbm is the practical number for quotation if automatic detection is a fundamental part of the intercept process, i.e., a signal power of −75 dbm has to be established at the receiver terminals for the initiation of any recognition action or control function by the receiver. Were it possible, as an alternative, to utilize an observer able to respond on a tangential signal-level basis, the receiver could be operated usefully with correspondingly lower input signal amplitudes.

The problems of operating automatic detection circuitry on an amplitude threshold basis revolve around the probability of a random noise peak exceeding a threshold amplitude (References 7 and 8). Figure 9-13 relates to this;
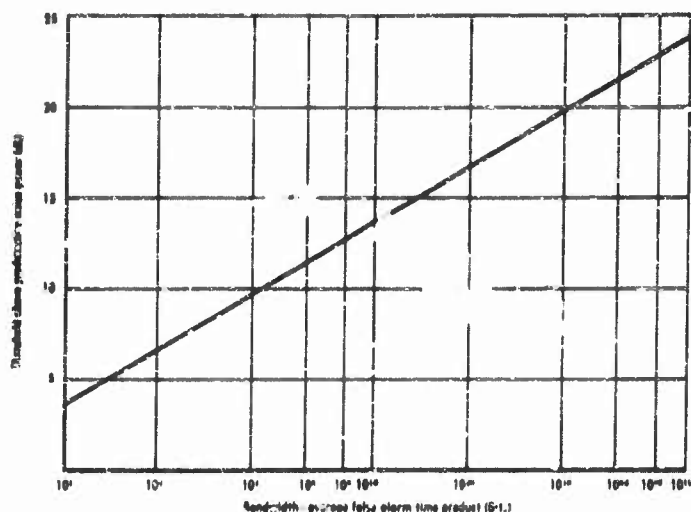
FIGURE 9-13   False alarms in electrical detection as controlled by threshhold setting.

a normalized quantity—bandwidth multiplied by average time between noise triggerings $(B \cdot t_a)$—is plotted against threshold setting in decibels above predetection average noise power. Although it is convenient to assume gaussian noise in signal versus noise problems, it should be recognized that this is an idealized circumstance not always encountered in practice.

Observation time and receiver effective noise bandwidth affect the interpretation of the data in Figure 9-13; the greater the bandwidth, the more rapid can be the rate-of-change of amplitude of a voltage or current in a system, i.e., a greater number of fluctuations can take place in a given observation time interval. Thus, a million noise peaks per second might be observed in a system having an effective noise bandwidth, B, of 1 megacycle. The probability of a noise peak exceeding an amplitude level 9 decibels above the average noise level is such that (from Figure 9-13) $B \cdot t_a = 5000$. The number of false triggerings per second is given by $1/t_a$. Thus, 200 false triggerings (and, possibly, false alarms) from random noise pulses could be expected in each second. With the same triggering level, the average false alarms would drop to two per each 10-second interval if the system noise bandwidth were reduced to 1000 cycles. Alternatively, if 200 false triggerings per second were considered excessive, an increase in triggering amplitude

threshold of only 5 decibels (with a corresponding decrease in effective
receiver triggering sensitivity of only 5 decibels to 14 decibels above average
noise) would reduce the number on an average to approximately one per
day. In other words, triggering reliability increases rapidly for very little
loss in sensitivity once a certain triggering level range has been reached; this
is of substantial operational importance.

The signal-to-noise power relationship plotted in Figure 9-13 can be inter-
preted in another manner. The threshold circuitry is confronted with a
combination of signal and noise. If the signal-to-noise ratio in decibels equals
the threshold setting in decibels above predetection noise power, there is a
50 percent probability that the combination would exceed the threshold set-
ting and that the signal would therefore be detected.

Detector characteristic and the relative pre-and post-detection bandwidths
affect the relationships shown in Figure 9-13 on a fractional-decibel basis
over common ranges of the bandwidth parameters.

From the foregoing, it is apparent that there is a substantial cost in system
sensitivity if electrical detection is used. This is true when the signal-to-noise
decision must be made on a purely amplitude basis. Unfortunately, this may
be the only basis if the search is for unknown signals of undefined charac-
teristics. When more is known of the characteristics of the signal sought—
pulsewidth, prf, pulse grouping, spectrum details, etc.—much of this loss in
sensitivity can be regained by the use of more complex decision circuits
tailored to those characteristics of the signal which differentiate it from noise.
In some instances, signals whose amplitudes are less than the average am-
bient noise can be selected reliably.

9.4.2 Received Signal Strength

It is a normal objective in receiver design to reduce the noise in the re-
ceiver. When this has been carried out as far as is practical, there still re-
mains the question as to whether or not a signal can be detected. Will the
signal power exceed that minimum value defining the receiver sensitivity?
This question focuses attention on the ranges of signal levels likely to be en-
countered in practical circumstances.

An intercept will be accomplished successfully only if the signal level at
the receiver input terminals equals or exceeds the threshold value necessary
to overcome successfully the total receiver noise power, N'_p, referred to the
input circuit.* The signal power actually delivered to the receiver under

*This is the power of a noise generator at the input terminals of a noiseless receiver of
like total power gain needed to account for the actual noise power in the output of the
actual receiver.

typical line-of-sight conditions will depend on the transmitted signal power, the range, the gain characteristics of the transmitting and receiving antennas effective over the transmission path, and propagation conditions. Presuming standard line-of-sight transmission, the received signal power is given by

$$P_R = P_T \quad \cdot \frac{G_T \, G_R \, \lambda^2}{(4\pi)^2 \, R^2} \tag{9-1}$$

where P_R = received signal power in watts

P_T = transmitted power in watts

G_T = transmitting antenna gain*

G_R = receiving antenna gain*

λ = wavelength in the same units representing range R

R = separation between transmitter and receiver

*The antenna gain values are effective gains applying *at the moment* over the transmission path. An isotropic antenna gain is taken as unity. An effective radar antenna gain might be 10,000 to 1 (40 decibels) in the main beam; but it also might be 1/10 (-10 decibels) if average back-lobe or zenith radiation conditions were applicable.

A useful alternate expression is given by Eq (9-3) which is developed by noting that the capture area A of the receiving antenna is related to receiving antenna gain and wavelength by

$$A = G\lambda^2/4\pi \tag{9-2}$$

$$P_R = P_T G_T A \, k'/4\pi \, R^2 \tag{9-3}$$

where A is the antenna cross-sectional area in the units representing range R, and k' is an "efficiency factor," commonly 0.6 for microwave parabolic antennas. The remainder of the units are as above.

It is possible to facilitate estimation of received signal power by use of the line-of-sight signal strength chart reproduced as Figure 9-14. The chart is actually two graphs plotted on one sheet. The *right vertical* and *upper horizontal* scales relate signal power density and distance between transmitter and receiver. For this calculation, the diagonal lines represent the effective radiated power (ERP) of the transmitter. The *left vertical* and *lower horizontal* scales relate receiving antenna capture area and operating frequency. For this determination, the diagonal lines represent receiving antenna gain over an isotropic antenna. The two plots have been combined to minimize the number of charts required to carry through a calculation. Each calculation is ordinarily a two-step process. For example:
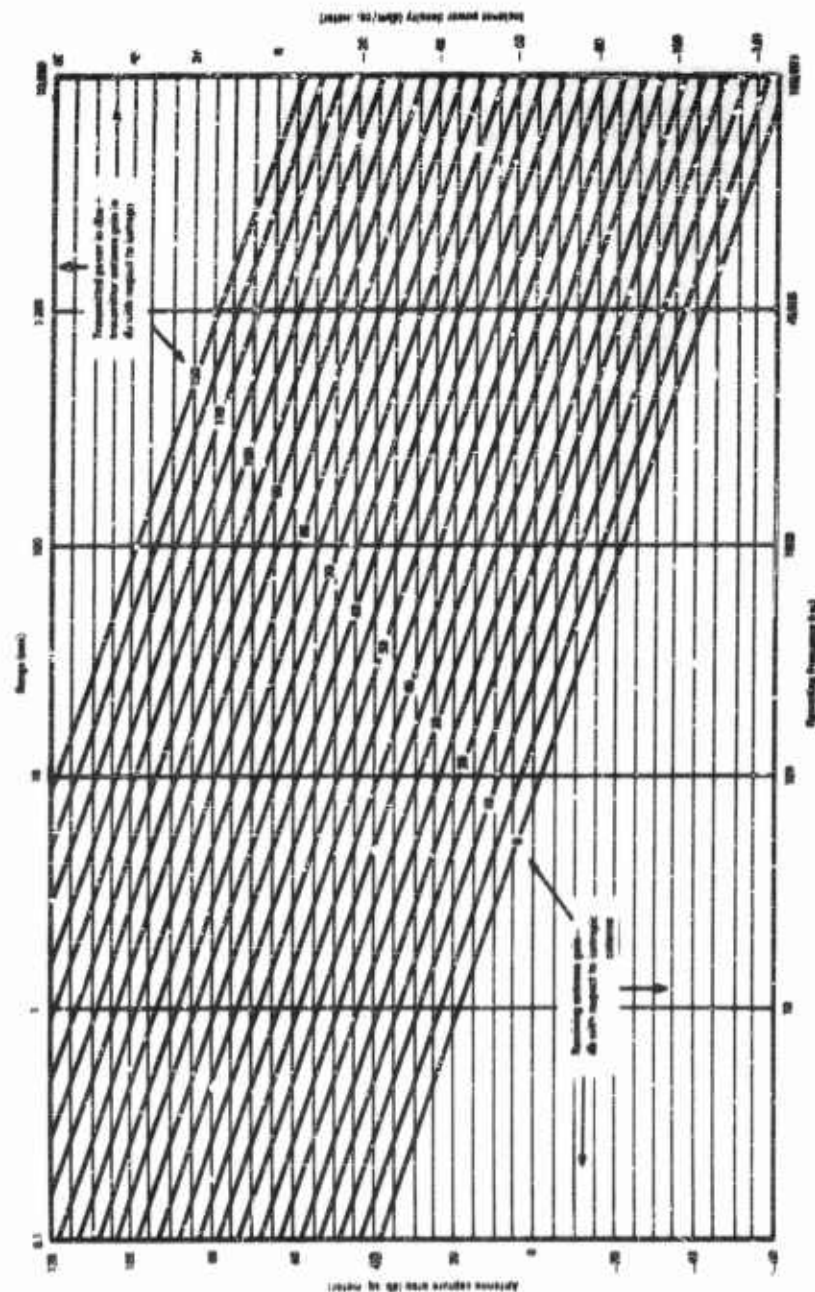
FIGURE 9-14 Frequency-power-range-antenna capture area nomograph.

TO DETERMINE POWER DENSITY

Given: Transmitted Power   = 100 kilowatts (kw)
        Transmitted Antenna Gain   =   35 decibels (db)
        Transmitter-Receiver Range   = 100 nautical miles (nmi)

First, determine the ERP for the transmitter in decibels above a watt (dbw). In the example, 100 kw = 10^5 watts, or 50 dbw. For an antenna gain of 35 db, the ERP = 50 dbw + 35 db = 85 dbw. Enter the upper horizontal scale at 100 nmi and project downward to the "85" diagonal. Project to the right vertical scale and read −2 dbm/sq meter as the power density of the receiver.

TO DETERMINE RECEIVED SIGNAL POWER

Given: Operating Frequency   = 10,000 megacycles per second (mc)
        Receiving Antenna Gain =   10 decibels (db)

Enter the lower horizontal scale at 10,000 mc and project up to the "10" diagonal. Project to the left vertical scale and read capture area as −31 db with respect to a square meter. The signal power at the receiver input terminal is then −33 dbm. This results from combining the values for power density (−2 dbm) and capture area (−31 db).

This example represents only one of several ways in which the chart may be used. It is possible to begin with a known receiver sensitivity and receiving antenna gain, and to determine the maximum range at which a given signal could be received. More generally, there are six variables involved: transmitter power, transmitter antenna gain, frequency, receiving antenna gain, receiver sensitivity, and range. Any one quantity can be solved for when the other five are given or can be estimated.

A very wide range of received signal powers can be anticipated under normal intercept conditions when the great variety of operational uses for a receiver are considered. This is illustrated in Table 9-I where some representative cases involving a typical range of transmitter powers, frequencies, antenna gains, etc., are tabulated.

An examination of Table 9-I indicates that very strong signal levels can be anticipated in many practical circumstances. This is indeed true and supports the use of relatively insensitive receivers (direct-detection receive. for example) where there is a resulting profitable improvement in ease of operation, technical simplicity, reliability, etc. It is sometimes argued, too, that additional sensitivity would serve only to add redundant or routine information of no great value and with the decided complicating aspect of increased signal density. However, this general argument must be considered with care since a somewhat more sensitive receiver can be operated on a less-sensitive basis (by an increase in acceptable threshold levels, for in-

TABLE 9-I TYPICAL RECEIVED SIGNAL POWER IN RADAR INTERCEPT

(Receiving Antenna Presumed To Have Unity Gain)
(Line-of-Sight Transmission)

Frequency (mc)	Transmitter peak power (watts)	Transmitter antenna gain (db)	Range (nmi)	Received power (—dbm)
1,000	10×10^6	40	100	—7
3,000	2×10^6	30	5	+2
			200	—30
			2000	—50
3,000	2×10^6	—10*	5	—38
			300	—73
			3000	—93
10,000	1	30	3	—72
		—10*	300	—145
10,000	0.25×10^6	30	10	—24
			200	—50
10,000	0.1×10^6	—10*	3	—63
			300	—98
40,000	0.1×10^6	55	10	—36

*Minor-lobe radiation.

stance). The amount of data obtained would be equal to that received from the basically less-sensitive equipment, but that portion received near the assigned threshold amplitude level will be cleaner (with respect to signal-to-noise ratio) and therefore possibly subject to a more meaningful analysis.

When maximum system sensitivity must be achieved, it is important to note the inclusion of receiving antenna gain in Eq (9-3). If maximum sensitivity is not the prime requirement, antenna gain may still be an important system parameter. For example, the improved system sensitivity provided by a directional antenna might permit a trade, in the receiver, of basic receiver sensitivity for receiver bandwidth, thus giving a greater intercept probability in some circumstances. Further, the directional characteristics (narrower beamwidth) accompanying higher antenna gain may be important in enhancing the reception of signals from a preferred direction, in excluding signals from certain directions, or in pinpointing the direction of arrival of selected signals.

The attention in this chapter to microwave receivers accounts for the emphasis on received power; it being convenient with microwave calculations to work in terms of power. At lower frequencies, precedent for expressing receiver sensitivity in microvolts has been set; the sensitivity is defined

as a signal voltage which establishes a suitable signal-to-noise voltage ratio. In that case, it is natural to consider the strength, in microvolts per meter, of the electric field established at the receiver antenna. This field intensity when multiplied by the effective height of the antenna provides the desired signal voltage value at the receiver terminals. These quantities are standard for low-frequency usage and need no further elaboration. Unfortunately, complex propagation conditions may well apply at the lower frequencies and no simple computation of signal strength [the counterpart of Eq (9-3)] ordinarily can be employed.

It is possible, of course, to convert an available signal power from a signal source to an equivalent voltage across the receiver input terminals when the circuit impedance conditions are specified. In this connection, impedance matching is a standard practice to be anticipated in most circumstances. Some exceptions where an improved performance can be obtained by some deliberate impedance mismatch between antenna and receiver are covered in Section 4.3 of Reference 1.

9.5 The Important Receiver Characteristics

There are substantial differences in intercept receiver design parameters with respect to those associated with receivers in general. Yet there are extensive basic similarities; a knowledge of receiver techniques in general, which this chapter presumes, provides a suitable framework so that in the following discussion attention can be given primarily to the differences in intercept receiver design characteristics.

As a preliminary, intercept receiver bandwidths and tuning processes are worthy of particular attention. Such consideration is facilitated by a general definition of certain related parameters.

D represents the total r-f bandwidth (usually in megacycles) assigned to a receiver for monitoring.

α is the incremental acceptance bandwidth of the receiver, i.e., the range of frequencies (a fraction of D) over which the receiver is sensitive at any instant. It is sometimes identified as the predetection bandwidth.

β is the postdetection bandwidth. It is the video (or audio) bandwidth.

B is the effective noise bandwidth and ordinarily will have a numerical value between α and β. The value will depend on α, β, the detector characteristic (linear, square law), etc.

t_s is the cycling time of the tuning process, if any. The incremental bandwidth, α, can be tuned across the total range, D, in the time t_s.

f_s is the "scan frequency" (presuming a repetitive tuning process). $f_s = 1/t_s$ cycles per second while the tuning rate, typically in megacycles per second per second, is given by D/t_s or $D \cdot f_s$.

Three other times are of importance in a discussion of receiver tuning processes.

t_d is the duration of a signal as defined by the time interval over which it is of detectable amplitude at the receiver. It may be very short—a few microseconds—with certain flash signals. It may be moderately short—a substantial fraction of a second—if amplitude conditions are satisfied when the transmitter antenna main beam is directed at the receiver. (The time then is a function of transmitting antenna rotation rate and the angular beamwidth through which a suitable signal strength exists.) It may be very long if the signal is of such amplitude that it can be detected via minor-lobe radiation from the transmitting antenna (the signal then is essentially continuous).

t_p is the interval between pulses of a pulsed transmission (often taken as the *longest* interval, i.e., as established by the lowest prf anticipated for any signal in the frequency range, D).

t_w is the pulsewidth of a pulsed transmission (often taken as the *shortest* pulse anticipated for any signal in the frequency range, D).

9.5.1 Receiver Noise: Noise Figure

The inevitable competition between signals and noise arises in intercept receivers. The basic problems in recognizing signals in the presence of noise were introduced in Section 9.4.1. There are difficulties, too, in measuring the characteristics of signals when their form has been corrupted by the presence of noise.

Johnson noise associated with input circuit resistance, with the resistive component of signal source impedance, etc., sets a minimum value to noise power evaluated at the receiver input terminals (Reference 9). This is the noise power, N_p, given by

$$N_p = kTB \qquad (9-4)$$

where N_p is in watts

k is Boltzmann's constant, 1.3×10^{-28} joule per degrees Kelvin (°K)

B is the effective noise bandwidth of the system in cycles

T is the temperature of the noise source in degrees Kelvin (°K).

If a temperature of 290°K were assumed (and this is generally applicable), and if the effective noise bandwidth of the receiver were one megacycle, the noise power inevitably associated with the signal at the receiver input would be 4×10^{-15} watts or -114 dbm. If there were no internal noise generated and if the nature of the system were such that a signal

having a peak power equal to this average noise power were recognizable as
a signal, it would be practical to quote the receiver sensitivity as −114
dbm on an equal signal-to-noise power basis.

The temperature T in Eq (9-4) refers to the noise temperature in degrees
Kelvin of the principal sources of noise affecting the system. An important
source may well be the input circuit of the receiver. If a 100-ohm resistor
bridged across the input terminals determined the input impedance, its tem-
perature, normally taken as 290° K (17°C), would apply.

If the 100-ohm physical resistance could be completely replaced by an
antenna having some equivalent antenna impedance (with a "radiation re-
sistance" component), the "antenna temperature" would prevail. (It is as-
sumed here t'   by ohmic losses of the antenna and its transmission system
are negligible— orarily true in intercept systems.) The temperature value
would depend on the temperature of the region surveyed by the antenna (to
which it is "coupled"). A directional antenna directed toward outer space
might have a very low temperature of a few degrees Kelvin. A redirection of
the antenna to another region of space might locate a "hotter" region, the
effective temperature would then be greater and the consequent increase in
external noise input to the receiver might readily be noted (as in radio as-
tronomy). If the antenna were again reoriented so that the main beam (and/or
minor lobes) intercepted an increasing sample of the "hot" earth, the an-
tenna temperature would increase rapidly and approach, again, the typical
290°K value. This is of particular import when considering the future role
in intercept receivers of very low noise r-f amplifiers—masers and para-
metric amplifiers. The potential reductions in system noise levels may be
very small if, as is true in many practical cases, external noise sources set a
minimum system noise only a few decibels below values established by
present receiver noise-figures (Reference 10).

To this point the discussion has been of ideal or near ideal receivers in
the sense that the dominating noise has been ascribed to a single identifiable
source effective at the input terminals of the receiver. Usually the receiver
circuitry itself will provide an additional, perhaps predominating, source of
noise that must be added to the irreducible minimum (kTB) in practical
calculations. The influence of this contribution is measured by the noise-
figure. Thus,

$$F = \frac{S/kTB}{S_o/N_o} \tag{9-5}$$

where F = noise figure of the receiver

S = available signal power from a signal source

S_o = available signal power at the receiver output

N_o = available noise power at the receiver output

k, T, and B are as defined before (and describe an irreducible minimum noise power associated with the signal, S).

Equation (9-5) can be rewritten

$$F = \frac{S}{S_o} \frac{N_o}{kTB} = \frac{N_o}{W} \frac{1}{kTB} \qquad (9\text{-}6)$$

where W is the power gain of the receiver for the signal (and presumably for input noise).

Thus N_o/W is the effective system noise level referred to the input of the receiver. The ratio of this value to kTB is a number greater than unity (it includes an equivalent noise ascribable to the receiver circuits) and is the noise-figure. For computational purposes, e.g., as in Eq (9-7), noise-figure is inserted as a number—the noise power ratio. However, the power ratio can be expressed in decibels, and a receiver is sometimes said to have an "eight-decibel noise-figure," etc.

A standard expression of computational value is that for the noise figure of two linear networks in cascade when both may contribute significantly to the total noise

$$F_T = F_1 + (F_2 - 1/W_1) \qquad (9\text{-}7)$$

where F_1 = noise-figure of Network 1

   F_2 = noise-figure of Network 2 in the circumstance that there is an
          impedance match between it and its signal source.

   W_1 = available power gain of Network 1

In this respect, intercept receivers are, of course, identical with standard receivers. Thus reference to a basic receiver text will produce valid information on the computation of noise-figures, the effects of cascaded networks, etc. (For example, see Section 1.1 of Reference 1).

However, some noise problems are emphasized in intercept receivers. One relates to the successful development of a low receiver noise-figure over wide r-f bandwidths. Recent work in low-noise amplifiers—TWT, masers, parametric amplifiers, etc., is having a major effect on this receiver design problem. R-F preamplification is now possible via amplifiers having fractional-decibel noise-figures and with sufficient gain (W) to guarantee that the noise-figure of the input amplifier will set that of the receiver as a whole. (Note in Eq (9-7) that a large value for W_1 will guarantee $F_T \sim F_1$.) Unfortunately, problems in bandwidths, tunability, and a lack of convenient

physical form currently plague those amplifiers having the more spectacular low-noise figures.

With the noise-figure concept in mind, it is possible to rewrite Eq (9-4) such that the quantity N'_p (equal to N_o/W) becomes the reference noise power (referred to the input circuit) of an actual receiver whose noise figure is F.

$$N'_p = kTBF \qquad (9\text{-}8)$$

In a practical receiver, it is the noise power N'_p with which an input signal must compete.

As an alternative to the noise-figure concept, the fact that N'_p is greater than N_p ($= kTB$) as defined in Eq (9-4) can be accounted for by assigning an elevated value to T. The results are equivalent; but it is sometimes convenient, particularly with low-noise amplifiers, to speak of a "20°K" amplifier. The meaning is that the noise situation can be valued by adding 20 degrees to T in the noise power calculation (and by assuming F to be unity). The noise contribution of this amplifier (a "20-degree" value) would be negligible in a typical situation (where it was to be added to an applicable temperature of 290°K for an external input noise source; it could be very important in a radio astronomy application if the equivalent antenna temperature were a few degrees. In other words, the contribution of the very low-noise amplifier is very important percentage-wise in radio astronomy; it could be very important in space age reconnaissance; it may be quite unimportant in some current intercept receiver applications.

A noise-figure of 10—good by standard broadband microwave receiver techniques—indicates the internal receiver noise contribution to have raised the total noise power at the receiver output by a factor of 10 over that which would have been measured with an ideal receiver. Accordingly, the sensitivity of the receiver would be reduced directly by the same factor (by 10 decibels); a -75 dbm sensitivity would drop to -65 dbm. A signal 10 decibels stronger than before would be required to compete on an equivalent basis with the increased noise. A major objective in receiver design, then, is to build receivers in which the internal noise contribution is kept to a value consistent with the sensitivity requirements.

The noise of concern to this point is largely ascribable to the receiver itself; it sets a maximum sensitivity (minimum noise) for a given receiver. In the lower frequency ranges of interest for intercept receiver use, a second type of noise will usually take precedence even over the total noise of the "imperfect" receiver. Included are contributions from natural static (lightning strokes, etc.), ignition noise, radio interference, etc., which increase

substantially in magnitude as the frequency decreases. The combination may well set a noise level substantially above the basic receiver value. This suggests that there may be little profit in improving receiver design (as regards internal noise generation for the radio frequency range in question, beyond the level which allows the external noise to predominate. Depending on the abilities of the receiver circuitry, the physical location of the receiver with respect to external noise sources, etc., the division point is usually in the 100 to 500 megacycle range, i.e., above 500 megacycles the internal system noise will set the basic system sensitivity.

As implied in Eq (9-4), the wider the effective noise bandwidth, B, the greater will be the receiver noise power with which the signal must contend. However, B is not necessarily the inherent bandwidth associated with the principal noise source (for which a value of T is established). It often is determined by a system bandwidth value following, and presumably constricting, that of the principal source of noise. Thus, the bandwidth of an antenna might be hundreds of megacycles in the sense of its ability to extract energy over that range from an incident wavefront. If the antenna were to feed a superheterodyne receiver, the effective system bandwidth, B, ordinarily would not be greater than the i-f bandwidth of the receiver.

9.5.2 Bandwidth Consideration: D, α, β, B

A consideration of intercept receiver characteristics reveals several important "bandwidths" as earlier defined. There is the r-f bandwidth D which describes the breadth of the total r-f range over which the receiver is capable of being operated. It defines the maximum band that can be assigned to a receiver for monitoring. Typically, the r-f bandwidth does not exceed an octave (a 2 to 1 frequency spread) although a receiver may include several tuning units (heads) each capable of covering separate octaves. They are ordinarily used alternately and D is set by the characteristics of whichever one is in use.

There is the acceptance bandwidth α which may or may not coincide in numerical value with the r-f bandwidth. It is the bandwidth over which the receiver is instantaneously sensitive. In a wide-open receiver the acceptance bandwidth corresponds to the r-f bandwidth since at any instant the receiver is equally responsive to signals anywhere in the r-f bandwidth; then $\alpha = D$. In other receivers, the acceptance bandwidth is less than the r-f bandwidth; α ordinarily equals the i-f bandwidth in a superheterodyne, and equals the effective bandwidth of the r-f amplifier in a TRF system. It is sometimes spoken of as the predetection bandwidth and is of prime importance in controlling the intercept probability properties of a receiver.

This predetection bandwidth bears a direct relationship to the common

receiver characteristics of selectivity and resolution. The ability to select one signal from a group of signals on a frequency difference basis, or to resolve two signals adjacent in frequency is set in an intercept receiver by the value of α in a completely normal manner.

Of particular importance to intercept receivers is "skirt" selectivity or "off-frequency" rejection. These terms refer to the response to signals far from the frequency to which the receiver is nominally tuned—the desirable goal is to minimize any such response. The problem is illustrated by the
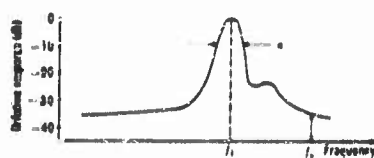
FIGURE 9-15   Selectivity curve, single-stage TWT TRF receiver. An off-frequency (f_2) signal is transmitted with only moderate attenuation.

typical response display curve in Figure 9-15 for a swept single-element TRF receiver. An acceptance bandwith in decibels down could be defined as α. But the off-frequency response in Figure 9-15 is down only 35 decibels; an off-frequency (f_2) signal exceeding by more than 35 decibels the minimum detectable level for an on-frequency (f_1) signal will produce a response in the output circuits that may well be interpreted erroneously as an indication of a weak signal on f_1. Fortunately, much can be done to insure acceptable off-frequency rejection by standard engineering techniques (via cascaded TRF stages or the superheterodyne i-f amplifier principle, for example).

The video or postdetection bandwidth β usually represents a design compromise influenced by requirements peculiar to intercept receivers. The selection must be consistent with the most severe requirement imposed by the need to reproduce to some degree the waveshape of any signal or class of signals anticipated for reception. Of obvious importance is the fidelity desired of signal reproduction. Wide bandwidths and appropriate attention to phase characteristics are necessary if there is to be a faithful reproduction of video envelope detail—leading edges of pulses, tilts, etc. If the video bandwidth is too small, a narrow pulse will not reach full amplitude. But if the principal objective is only signal detection, a considerable reduction in video bandwidth is allowable for only a small loss in weak-signal detectability since there is a simultaneous reduction of noise in the output circuitry. Because of this, a broad maximum results in the curves representing signal detectability versus video bandwidth (for different predetection bandwidths) under typical conditions (Figure 9-16, and Figure 8.30 of Reference 11).

If only c-w or audio amplitude-modulated signals were anticipated, β could be narrowed by orders of magnitude (with profit in reduced noise). In view of this, it is not uncommon to find more than one postdetection amplifier paralleled in a general purpose receiver. A wideband amplifier (many
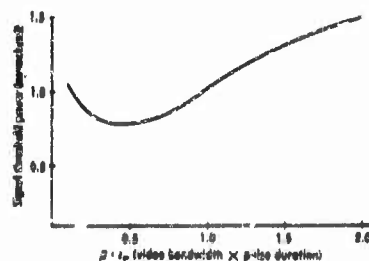
FIGURE 9-16 Signal threshold power of a crystal-video receiver as a function of video bandwidth.

megacycles in width) might be associated with a high-fidelity "analysis" output channel, a circuit of more moderate bandwidth (leading to a better signal-to-noise ratio) might be available for regular pulse-signal detection and display; and a third, very narrow channel favoring a high signal-to-noise ratio for c-w signals might be included as an additional option.

It is often true that α may necessarily exceed substantially the value required to pass the frequencies present in the spectrum of the intercepted signals. This contributes to a value of effective noise bandwidth B in excess of that which would be optimum from purely maximum sensitivity considerations; this accounts in part for why an intercept receiver may be less sensitive than a comparable narrow band, specialized (radar) receiver representing an equivalent state-of-the-art.

Actually, B, as used in Eq (9-4) and (9-6), is a complicated function of pre- and postdetection bandwidths, signal level, the type of signal (c-w or pulsed), and detector characteristic. With respect to the last item, a common situation of particular practical importance is that in which the signal and noise powers are roughly equal. This often occurs as a true "weak-signal" condition in which the detector exhibits square-law characteristics. If signal and noise powers at the detector are of comparable magnitude, a square-law characteristic is usually applicable regardless of absolute power level and independent of the basic detector characteristic at that level for a noise-free signal (pages 301-308 of Reference 8).

Presuming a square-law characteristic, B becomes dependent on both pre- and postdetection bandwidth (α and β) as represented in Eq (9-9) which is accurate for pulsed signals (Chapter V of Reference 12).

$$B = (2\alpha\beta + 3\beta^2)^{\frac{1}{2}} \pm 2\beta \qquad (9-9)$$

Equation (9-9) presumes α to be at least as large as β; that is the usual circumstance. Should β exceed α, it is appropriate to assume in the calculations that β is equal to α. The selection of sign for the second term ($\pm 2\beta$) depends on whether or not a contribution in the output brought about by a product of noise and signal pulse is considered signal (to decrease B) or noise (to increase B). It is sometimes ignored; this is not safe when α and β are comparable in magnitude because the effect on B is substantial. This con-

tribution has an important practical effect because this signal-dependent noise superimposed on the pulse form can distort the pulse shape and thereby affect triggering circuitry. In the special case, particularly important to intercept receivers, where α is substantially larger than β, Eq (9-9) simplifies to

$$B = (2 \alpha \beta)^{\frac{1}{2}} \qquad\qquad (9\text{-}10)$$

A rectangular passband is implied with the use of an α value. This is substantially satisfied in practice with multistage filter circuits. If α, however, is set by the 3-decibel bandwidth of a single tuned circuit, the effective bandwidth for noise is 1.57α; it is 1.22α for two such stages in cascade.

9.5.3 Gain and Dynamic Range

As is true for receivers in general, maximum gain requirements are set by the need for the ability to raise the level of the input signal of minimum usable amplitude (as determined by signal-to-noise considerations) to a specified minimum output amplitude value. A maximum gain of one hundred decibels would not be an unusual requirement. A complication is injected in many intercept receiver applications by the need to handle signals of widely varying amplitude in rapid sequence—in a time framework permitting no manual adjustment of a gain control. Yet it is generally desirable to retain some substantial representation of original amplitude variation in the receiver output signals so as to permit modulation determination, measurement of antenna scan characteristics, relative signal amplitude determination, etc. The result is the imposition of severe requirements on dynamic-range characteristics. While some compression of signal-level variation in the output must be anticipated (in view of a possible 100-decibel range in input signal levels), it is important that the compression (clipping, limiting, etc.) be of an acceptable character. This has led to much attention to automatic gain control (agc) circuitry, amplifiers with logarithmic characteristics, etc. It is a factor affecting the gain distribution throughout a receiver—r-f gain (if possible) versus i-f gain versus video amplifier gain, etc., there being advantages (in recovery time, for example) in allocating certain gain-control functions to certain amplifier types. The subject is of such importance as to justify special treatment in Chapter 24. A particularly difficult problem may exist in the simple direct-detection receiver where a square-law detector characteristic can increase the dynamic-range problem in the following video amplifier. (A 10-db amplitude change before detection may translate into a 20-db change in the video amplifier.) In that receiver, the video amplifier provides the only opportunity for control of gain characteristics. The situa-

tion in a simple crystal-video receiver is relieved somewhat by its lower sensitivity, with a correspondingly smaller dynamic range required in normal operation.

9.5.4 Tuning Range and Tunability

The tuning range of an intercept receiver is defined as that total frequency excursion (the r-f bandwidth D) over which the receiver acceptance bandwidth can be adjusted, i.e., "tuned". When the acceptance bandwidth, α, matches the r-f bandwidth (as in the direct-detection, "wide-open" receiver) there is no tuning to be done and "tuning range" has no meaning. But when the acceptance bandwidth is less than the total r-f bandwidth, a tuning process must be instituted if the total r-f bandwidth is to be monitored. The ease with which the tuning can be accomplished and the flexibility in the tuning program relates to "tunability".

Tunability is of prime importance in intercept receivers. In combination with acceptance bandwidth characteristics, it largely sets the intercept probability of the system as regards search for signals in frequency. This importance has generated a wide variation in tuning techniques and associated features. In terms of sophistication (and this is not necessarily a measure of value) tuning processes range from mechanically operated (by hand or motor drive) "slow" sweeps measured in seconds per scan, through electronic tuning of many scans per second (rapid sweep), to situations where tuning rates are measured in thousands of megacycles per second per microsecond (Microsweep). The term "scan frequency" ordinarily refers to the number of periodic sweeps across the r-f band (or subband) per second. The "scan rate" is the rate of change of frequency (often in megacycles per second per second).

The mechanisms for accomplishing this variety of tuning schemes are many and reflect the latest innovations in components and circuits—electronically tuned oscillators and amplifiers, dielectric tuning, permeability tuning, etc. The important components and techniques are described individually in Chapters 26, 27, and 28.

While the necessary developments of flexibility in tuning continually tax the state-of-the-art, intercept receiver requirements in tuning stability and in automatic frequency control (afc) are often not severe. This reflects the fact that there is usually only a short-term interest in individual signals. Frequency measurement is apt to be important on a relative (rather than on an absolute) basis. There are exceptions in uses that arise in the lower frequency ranges, or if a precise analysis of a stable microwave signal is to be conducted, or in correlating measurements made with different receivers. The requirements in stability and frequency control can usually be met by

standard receiver techniques, but there is often an associated problem aris-
ing from the need to stabilize unusual tuning mechanisms that may ordinar-
ily be operating at high tuning rates.

Some of the properties of tunable receivers are uniquely linked to the
tuning rate. These can best be described using the time and bandwidth
definition included in the introduction to Section 9.5.

Thus, in a slow-scan receiver the relationship $\alpha << D$ usually holds.
Some important characteristics are set by the time required to tune through
D (the time t_s) with respect to the effective signal duration $t_d << t_s$
(as with flash signals), intercept probability will be very poor. If $t_d < t_s$
(where t_d, perhaps, is conditioned by the main-beam illumination time of
the receiver by the transmitter antenna), the intercept probability will be
poor. If $t_d > t_s$ (where t_d becomes long because of a possible ability of the
receiver to receive a "continuous" signal via minor-lobe radiation), the
intercept probability can be excellent.

It is often important in searching for pulsed signals that a narrow-
band scanning receiver not tune through the signal frequency in the
interval between pulses, t_p. The following considerations then apply. The
acceptance bandwidth, α, often is fixed (perhaps by resolution require-
ments); the reception time on-frequency is then set by α and the scan
time, t_s, and must as a minimum value be at least equal to t_p. The reception
time might better be set to a value $(n-1)t_p$ where n defines a number of
consecutive pulses as required to distinguish a signal (in contrast to a
single noise pulse) or to identify a signal by some combination of charac-
teristics. This on-frequency time is guaranteed by a minimum scan time,
t_s, such that*

$$t_s = (n-1)t_p D / \alpha \qquad (9\text{-}11)$$

The reciprocal relationships between scan frequency, f_s, and scan period, t_s,
and between prf and the interval between pulses, t_p, gives the expression

$$f_s \text{ (in cycles per second)} = \frac{(\text{prf})}{(n-1)} \cdot \frac{\alpha}{D} \qquad (9\text{-}12)$$

In a rapid-scan receiver, the scan time, t_s, is ordinarily made equal to or
less than the signal duration, t_d. When t_d is fixed by the main beam trans-

*There usually will be an uncertainty as to values to be selected for the signal charac-
teristics t_s, t_w, t_p ($= 1/\text{prf}$), etc. The worst conditions anticipated from the intercept
point-of-view for signals in the range D are often taken, i.e., shortest pulse, lowest prf,
etc.

mission time of a scanning radar antenna, the scan frequency, f_s, ordinarily will be set to a few tens of cycles per second. This rate is consistent, too, with the development of a flickerless panoramic visual presentation of signal activity in the monitored band. The total number of pulses transmitted by a radar in the interval t_c is given by the product of t_d and the prf. The number of pulses intercepted in the interval t_c on an average will be given by $M = \alpha$ prf t_d/D when typical numbers are inserted *for a short transmission interval, t_d*, it becomes obvious that α must be large (perhaps 10 percent of D) if a recognizable output signal is to be developed.

A *microsweep receiver* is one in which ordinarily $t_s \leq t_w$, i.e., the receiver scans the total range, D, in the time duration, t_w, of the shortest anticipated pulse. Intercept probability is again high though there are attendant problems in circuitry. The bandwidth, α, is carefully selected to achieve useful resolution characteristics and to minimize rise-time problems (considering the very short fractional pulses which must be dealt with in the receiver circuits).

Frequency scanning implies a time-sharing of the attention of the receiver. The design philosophy in scanning receivers is to arrange matters so that each signal receives sufficient immediate attention to establish quickly the information desired from it. The hope is that the lack of continuing attention will, in effect, exclude only largely redundant information. It is to be noted that over a long averaging period (over a long t_d or over many periodic, short t_d intervals), the intercepted fraction of the total signal available to the receiver is fixed only by the ratio α/D and is independent of scan rate. There are, however, optimum choices of α and scan rate which insure that some portion of the transmission will be received *early* in the transmission interval (for high intercept probability) and that the α/D fraction represents a form of the signal, perhaps groups of consecutive pulses, such that the intercept has meaning.

Generally speaking, the non-scan receiver has high intercept probability but essentially no resolution in frequency; sensitivity is usually low. The slow-scan receiver features high intercept probability for some signals, very poor intercept probability for others, good resolution in frequency and, ordinarily, high sensitivity. The rapid-scan and microsweep receivers feature generally good intercept probability but necessarily compromise their abilities in resolution and sensitivity depending on actual scan-rates and bandwidths. Thus, each receiver type has attractions and deficiences in basic intercept and signal-selection characteristics linked to tuning processes and tunability. These compromises appear, as well, in other aspects—stability, data-handling abilities, versatility with respect to the detection and analysis of unusual signals, etc. Each type has virtues that recommend it for certain important

uses, and much receiver design work has been concerned with maximizing the operational utility associated with a given type.

9.5.5 Spurious Signal Response

Some substantial attention must be accorded in receiver design to the possible internal generation of spurious signals and responses. Several sources will be mentioned later: images, harmonic mixing, etc., in superheterodynes; the generation of harmonics and cross-modulation products in r-f preamplifiers, direct-detection in mixers, direct-detector response to strong signals in another r-f band, etc. In another area, local-oscillator radiation can be troublesome (when received by a second receiver); and transient disturbances from intermittent switching operations, electrical interference from blower motors, etc., can be interpreted falsely as signals by the receiver circuitry.

There is another aspect to the spurious signal problem that is largely beyond the control of the receiver designer. There may be available to the receiver actual radiated signals that are false in the sense that they were not intentionally generated. Thus, physical objects local to the transmitter or receiver can generate "site" reflections that can add "pulses" to the general signal environment, or can distort the apparent characteristics of actual received pulses. Doppler shifts can be important. Transmitter moding can produce large-amplitude signals on frequencies not harmonically related to the basic transmitter frequency. Transmitter harmonics are common, and when transmitted beam powers are rated in billions of watts, a spurious signal down 40 or 50 decibels is still appreciable. Multipath propagation, tropospheric scatter transmission, etc., can distort signal shapes, modulations, apparent directions-of-arrival, etc. This false information cannot be charged to the receiver; it is truly contained in the radiations actually present. The important thing is to recognize the distortions for what they are, and to realize that an exaggerated refinement imposed on receiver measuring abilities may be fruitless and costly if the receiver must operate in an environment where the quality of the data available to it prohibits intelligent use of these abilities.

There is a distinction to be drawn between these distortions of intentionally radiated signals and the radiation of electromagnetic energy incidental to certain industrial or research processes. The detection of the latter class of signals may well be legitimate objectives of a reconnaissance operation because of the information imparted by such signals.

9.5.6 Data Read-Out

An important factor in the design or selection of a countermeasures receiver is its ability to work with associated equipment in accomplishing a

particular task—to work with d-f equipment, with a particular signal analyzer, or with a certain signal read-out unit. A receiver may have an extra utility because the form of the output data is readily adapted to the needs of a computer or telemetry equipment. A receiver may be an optimum choice for a particular task because of unique abilities to handle certain signal parameters, or combinations thereof; this, in turn, leading to desirable abilities in signal segregation and identification, or in new signal recognition (new in characteristics, or new in "time" in the sense of the recent appearance of the signal).

Thus, an important variable in intercept receiver design is the form, or forms, of data read-out. The variety stems from the variable natures of the intercepted signals, and particularly from the many uses to which the output data may be put. Consideration must be given to the form of data presentation for analysis, whether this process is to be accomplished at the time of intercept or later. Although these processes are discussed in detail in Chapter 11, their influence on intercept receiver design and operation is quite significant and should be recognized. Two basic approaches can be outlined: quantized and continuous. In the former, storage channels or electronic codes (of predetermined range) are used for characterizing each signal parameter, and it is necessary only to measure the parameter with sufficient accuracy to quantize it into the appropriate storage channel. This approach is more easily adapted to digitizing and subsequent machine sorting and processing of data, which may be the only feasible means of reducing the tremendous amounts of data intercepted in a high signal density environment. However, the "fine structure" of each signal is destroyed, and significant technical information may be lost. The continuous (or analog) approach alternatively involves presentation and/or storage of signal envelope information in a near-original form, cathode-ray tube (CRT) displays, motion pictures, strip films, direct tape recording, etc.). Analysis in this case usually requires substantial human involvement, even with sophisticated auxiliary read-out equipment. Further, the postdetection bandwidths required are usually greater. While the optimum output data forms suggested by the analysis needs often can be provided through suitable engineering design, certain types of receivers are basically more compatible with these needs. This is discussed further in Section 9.6.

9.6 Receiver Identification

It is convenient for discussion to group types of intercept receivers in accordance with some common electrical or operational characteristics. There are several bases for classification, the substantial number stemming from the several major factors that influence intercept receiver design.

For example, receivers are sometimes classified on the basis of frequency range—LF, HF, VHF, microwave, etc. Sometimes the types of signals sought provide the differentiation—communications, radar, missile guidance; or pulse, c-w, FM, etc.

A means for grouping is provided by the intended time-utilization of the intercepted data. The data may be utilized immediately to warn of a missile attack, of the presence of an ASW aircraft, etc. The data may be used in the near future—to aid planning of a new strike route, the selection of countermeasures equipment to be carried on the next mission, etc. Or the data may be utilized months from the time of intercept, following prolonged analysis to extract technical information of long-range importance. This time segregation carries with it implications in data quality and total processing since the demands on accuracy of measurement are substantially different in receivers for warning, strategic search, special search, etc.; so also are demands on intercept probability—the assurance that the important signals are being detected in the minimum possible time.

It is possible to catalog receivers by basic circuitry (direct-detection, superheterodyne, etc.). Another mechanism is furnished by circuit features of particular meaning to intercept receivers such as tuning (scanning) characteristics, acceptance bandwidth characteristics, etc. Cataloging in this chapter will follow the basic circuit differentiation because of the concern with techniques; there will be discussions of two basic types (1) *direct-detection* receivers, with and without r-f preamplification, and with contrasting acceptance bandwidths and tuning processes, and (2) *superheterodyne* receivers similarly differentiated. In addition, mention will be made of some *miscellaneous* types not conveniently catalogued as above—superregenerative, direct-display (spectrum analyzer), etc. There will be no attempt to cover all intercept receivers currently in use, in production, or in development. (For such information see Reference 3).

9.7 Direct-Detection Receivers

These receivers are characterized electrically by a direct conversion from the radio-frequency form of the signal to a video or audio frequency counterpart—the video- or audio-frequency modulation is recovered by direct detection of the radio frequency signal. The simple crystal-video receiver is the outstanding example (Figure 9-17a). As a class, these receivers exhibit relatively poor sensitivity and selectivity, but also simplicity, light weight, and low-power consumption. It is almost axiomatic that as the sensitivity or selectivity characteristics are improved by circuit innovation (with preamplification, for example) there is a corresponding loss in the attractive size and weight characteristics that are identified with the basic crystal-
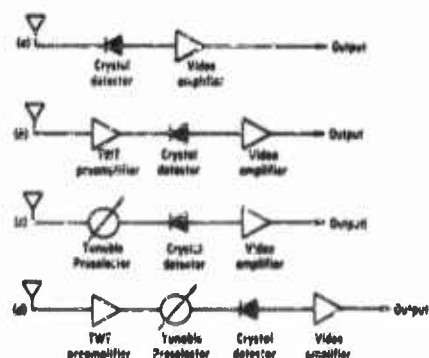
FIGURE 9-17   Direct-detection receivers.

video unit. However, the results provided by the more complex assemblage may still be very attractive and quite competitive with other receiver design approaches.

If no r-f preamplification is employed, the maximum acceptance bandwidth of the direct-detection receiver (and, hence, the total r-f band monitored) is limited by the frequency characteristics of the antenna, by any passive r-f filter that might be used, or by the r-f bandwidth of the detector unit. Operation is obviously restricted to the overlap frequency range of any band-limiting devices that may be interposed ahead of the detector element. In this circumstance, over-all receiver sensitivity is set by the characteristics of the detector unit and is substantially independent of acceptance bandwidth. (This presumes some care to insure low-noise properties in the video amplifier.) In other words, the receiver noise output is set by the noise properties of the detector, these being such that detector noise overrides the contribution from other circuit elements (Chapter 19 of Reference 1).

If r-f preamplification is employed, it is usual that the r-f bandwidth properties of the preamplifier will set the bandwidth characteristics for the receiver. Also, the sensitivity will be set by the acceptance bandwidth and noise figure of the preamplifier. (This presumes a certain minimum gain in the preamplifier, relative to its noise-figure and to the noise-figure of the following stage. The preamplifier noise contribution then predominates in the total receiver output noise; see Reference 13).

9.7.1   Broad Acceptance Band Direct-Detection Receivers

This category includes the untuned, wide-open, crystal-video receivers. The total r-f band monitored corresponds identically to the acceptance bandwidth; there is no tuning. The particular attractions are the simplicity, light weight, small size, low power consumption, etc.; and, for many uses, the high intercept probability. Intercept probability is unity *for those signals having sufficient amplitude to be detected and recognized*. Unfortunately, the sensitivity is apt to be low, tangential sensitivities of -45 to -55 dbm being typical in the microwave range below K-band and with reasonable (megacycle) video bandwidths. When "electrical" detection (automatic trig-

gering) is employed, a practical sensitivity range is -35 to -45 dbm with a reasonable false-alarm rate.

The heart of the receiver is the detector unit; both bandwidth and sensitivity are directly related to its properties. The substantial research and development efforts on detectors are producing continuing improvements. As of this time (circa 1959) r-f bandwidths from 2:1 to 10:1 are standard and tangential sensitivities of -55 dbm or better are anticipated (with a one-megacycle video bandwidth).

Detector units are under development for use to frequencies substantially above 100 kilomegacycles per second, and in many of the higher frequency bands this receiver type represents, at the moment, the only available technique.

The drawbacks are the low sensitivity, and the absence of frequency selectivity. Precise measurement of frequency is not possible. Because of the lack of resolution in frequency, data rates are high in high-signal density areas, and trouble may be encountered in establishing any practical individual signal recognition or selection.* Because of the wide variations in the amplitudes of incident signals simultaneously present there is apt to be a dynamic range problem. Ordinarily, manual gain adjustment is not feasible; the circuitry must be able to handle the many signals without gain adjustment. The gain control problem is complicated in the video circuitry because of the dynamic range expansion resulting from the square-law detector characteristic applicable over much of the r-f input signal amplitude range. Considerable attention to the design and use of logarithmic video amplifiers with this type of receiver has resulted. Transistor circuitry has a natural association with the small size, weight, and power consumption properties desired in these receivers and some truly remarkable units have been developed.

The physical nature of the r-f circuitry is dependent on the frequency band involved and is ordinarily, though not necessarily, of coaxial, stripline, or waveguide configuration. Because of the low sensitivity it is not uncommon to locate the detector unit immediately following a directional microwave antenna. There results a sensitivity benefit from the antenna gain and from the reduced transmission-line losses. Thus, this configuration provides a space discrimination (fixed or scanning) with a sensitivity improvement analogous to the frequency discrimination and sensitivity improvement obtained in a tunable receiver.

*However, the resultant wide-open nature of the receiver may well permit the effective measurement of other signal parameters, so as to completely offset the inability to measure frequency. Indeed, if a single emitter transmits programmed or random pulse-to-pulse frequency variations, a frequency-selective intercept receiver would be in some difficulty.

9.7.2 The Untuned Receiver With R-F Preamplification

A basic attack on the sensitivity problem is often made via broadband r-f preamplification with a TWT usually providing the basic mechanism. Bandwidth and gain characteristics then are controlled by the r-f amplifier tube as shown in Figure 9-17(b). A 2:1 r-f bandwidth is common up through X-band with narrower bandwidths possible in the lower K-band. Noise figures of 5 to 25 decibels are not uncommon, as seen in Chapters 26 and 27, depending on the nature of the TWT. The lower noise-figures are ordinarily associated with physically heavier tubes, narrower-band tubes, and, to some extent, with lower frequency tubes. Despite a higher noise-figure, the characteristics of the light-weight tubes (permanent magnet focussed, electrostatically focussed, etc.) are more consistent with the weight and power consumption objectives of the crystal-video receiver and are frequently chosen for these advantages (Figure 9-18).

The tangential sensitivity of broadband receivers with TWT preamplification typically range from -60 to -80 dbm in accordance with TWT noise characteristics. Any such sensitivity improvement accentuates tremendously the potential confusion arising in high-signal density environments. Fortunately, the nature of the TWT r-f preamplification affords a partial remedy through the automatic r-f gain control (agc) processes made possible. One such technique assists a sequential attention to signals on an amplitude basis in those instances when the scanning actions (in direction) of transmitting and, possibly, receiving antennas (plus the possible physical motion of transmitter and/or receiver) tend to cause each of the signals, in turn, to be the strongest present. The agc action can be made to favor the strong signal by causing a momentary suppression of the others present; the tendency, then, is to produce a clean "look" at each signal in turn. There results a "time segregation" of signals, which can be an important aid to signal selection.

9.7.3 Direct-Detection Receivers With Tunable Preselection

The problems in data handling occasioned by the "wide-open" operation of the broadband receiver can be largely relieved by deliberately restricting the acceptance bandwidth of the receiver. This can be done by adding a narrow-band, passive r-f filter adjustable in frequency (Figure 9-17c). The principal gain is in selectivity (and, hence, in reduction of interference) and in accuracy of frequency measurement. There is at best no gain in sensitivity; presuming no preamplification, the acceptance bandwidth variation has no practical effect on receiver noise output power. There is a loss in intercept probability since, basically, the attention of the receiver to an r-f signal frequency must be time-shared with the other frequency increments that together compose the total band to be monitored.

Important parameters in connection with a tunable, passive filter are its acceptance bandwidth (usually the 3-decibel bandwidth), the mid-band insertion loss, the "skirt" selectivity (a measure of the ability to reject large amplitude, off-frequency signals) and the tuning mechanism—manual or motor drive, tuning speed, etc. Ordinary filter theory applies, i.e., insertion loss increases but skirt selectivity improves as the number of cascaded circuits in the filter is increased. Bandwidths range typically from a fraction of a percent to several percent of the total tuning range. The design of tunable passive filters for preselector use is discussed in Chapter 28 of Reference 2.

There are but few practical receivers of this type. The addition of a tunable filter penalizes the basic simplicity of the crystal-video receiver while the remaining lack of sensitivity is a definite drawback. In the absence of substantial sensitivity (sufficient to permit reception of minor-lobe radiation from the transmitting antenna) there is a serious loss of intercept probability brought about by the reduction in acceptance bandwidth. (This is discussed in Section 9.5 where t_d will be at least moderately short and $t_d < t_s$.)

9.7.4 The Addition of Preselection Plus R-F Preamplification

The electrical performance objections to the direct-detection receiver plus a passive, tunable filter are largely removed when r-f preamplification is added to the system. Physically, this can be done in two ways: by adding a broadband untuned preamplifier such that the receiver bandwidth is still determined by the passive filter, or by adding a tunable preamplifier having, in itself, the acceptance bandwidth characteristics desired for the receiver.

If the broadband preamplifier is used, there are two possible configurations: it can be connected directly to the antenna with the narrower-band tunable filter interposed between it and the detector unit, or it can follow the filter and directly precede the detector. There are arguments for each scheme with the majority favoring the use of the preamplifier as the input element (Figure 9-17d). In particular, if the filter follows the amplifier it then restricts the noise bandwidth as well as the signal acceptance bandwidth and the sensitivity is conditioned by the preamplifier noise figure (presuming reasonable preamplifier gain) taken with the following narrow r-f filter bandwidth. Otherwise sensitivity would be penalized since the noise bandwidth with the filter preceding the amplifier would remain as the total bandwidth of the preamplifier even though the incremental bandwidth for signal reception α would be reduced. In addition, the preamplifier (as the input element) often can be operated physically at a location remote from the receiver itself, perhaps at the antenna, so that subsequent transmission cable or waveguide losses do not impair the basic sensitivity of the receiver

(since both noise and signal are attenuated equally). It is usually more difficult to operate a tunable filter by remote control as would be necessary were it to be associated with the remote antenna as the input element.

There is one important advantage in having the filter precede the amplifier. Limiting may well occur in the amplifier because the dynamic range of input signals may far exceed the amplifier capabilities. If two signals are present in the amplifier simultaneously, mixing is possible such that output signals on spurious frequencies can be produced; they may be interpreted as genuine when a filter (following the amplifier) tunes to a spurious frequency. If the filter precedes the amplifier, any signals passing through the filter must have been present at the receiver input and to that extent they can be classified as legitimate.

A preamplified video receiver with two 7-14-kmc channels is shown in Figure 9-18.
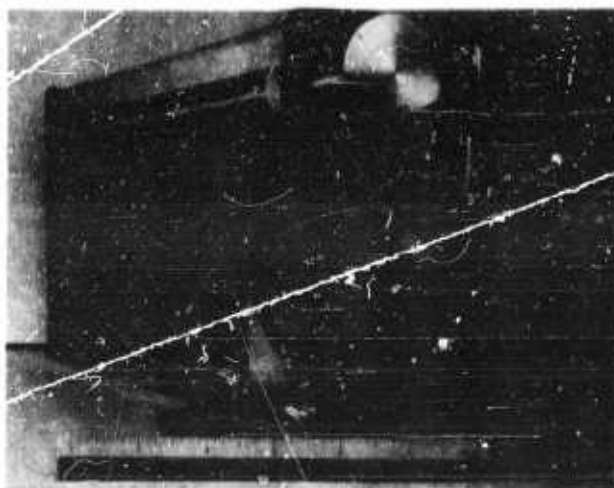
FIGURE 9-18 A preamplified crystal video receiver. This unit contains two separate 7-14 kilomegacycle channels, each having a TWT preamplifier, a detector, and a video amplifier, as well as an integral power supply.

The second important class of receivers of this basic type employs electronically tuned preamplifiers having in themselves the acceptance bandwidth characteristics desired for the system. Dispersive TWT amplifiers and backward-wave amplifiers, both voltage-tuned devices, represent the principal components for use in the microwave ranges. The dispersive TWT

is of marginal utility because its gain is variable with frequency; so too is its acceptance bandwidth. The latter quantity is quite large (usually 10 to 20 percent of the center radio frequency). Noise figures ordinarily are not good and the skirt selectivity is moderately poor. Even though some improvement in bandwidth characteristics can be obtained by cascading tubes (to form a multistage amplifier) the over-all characteristics are unattractive for many important uses. A sensitivity of −55 to −65 dbm at S-band is typical.

The backward-wave amplifier promises better performance in terms of noise figure, gain, and resolution—acceptance bandwidths of 0.1 to 1.0 percent being quite reasonable. Since the resulting receivers are electrically tuned, a great versatility in tuning is possible; this leads naturally to panoramic visual presentation of intercepted signals as a primary direct display mechanism; it provides for the automatic signal control of the frequency scan process (to put the receiver on-frequency); and it provides an output voltage related to frequency (the control for the electronically tunable element) that is convenient for recording and automatic processing purposes.

Basically, the tuned, narrow-band, direct-detection receiver with r-f preamplification—a true TRF configuration—constitutes a very useful receiver class. It is excellent competition for the superheterodyne as regards noise figure and bandwidth (and hence sensitivity), and tunability. In some instances it is less troubled with spurious responses (a very important aspect). S-band sensitivities in the −100 dbm range for a 5-megacycle acceptance bandwidth have been developed. However, the uniformity of characteristics as a function of frequency, the skirt selectivity, and the versatility (in bandwidth adjustment) are factors apt to be poorer than those found in a comparable superheterodyne. The superheterodyne now is ordinarily favored in those frequency ranges in which components for both exist. but advances in techniques are causing increased attention to the TRF receiver.

9.7.5 Multiple-Channel, Direct-Detection Receiver

The limited acceptance bandwidth of the TRF receiver necessitates a tuning program. But in no event will the proportion of time spent on any one frequency, on the average, exceed the ratio of acceptance bandwidth to total r-f bandwidth. Thus, the time devoted to any frequency may be very small and intercept probability will certainly suffer in some important circumstances. The "multiple-channel" direct-detection receiver is an ingenious approach to high intercept probability with the retention of frequency resolution. (The instantaneous frequency indicators of Section 9.10.1 have similar attractions. A multiple-channel receiver is described in detail in Reference 14).

Such a receiver is illustrated in Figure 9-19. A multiplicity of passive,
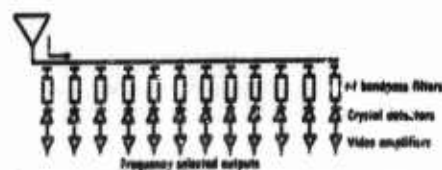
FIGURE 9-19 Multichannel direct-detection receiver.

fixed-tuned filters provides contiguous frequency coverage over a band given by the product $m\alpha$ where m is the number of filters and α is their individual acceptance bandwidth. The filters are coupled in parallel to the antenna and each feeds an individual detector-video amplifier combination. As a consequence, the receiver is always sensitive on all frequencies. Thus, it has high intercept probability for signals of sufficient strength regardless of frequency—but it has, as well, the resolving power and frequency measuring ability associated with the individual filter bandwidths. This is an attractive combination; a receiver sensitive to all frequencies at all times but with substantial resolution in frequency.*

In effect the receiver quantizes the intercepted information by discrete frequency subbands. A logical, simple visual display can be provided by a bank of lamps; an activated light signifies occupancy of the incremental bandwidth with which the lamp is associated. Digitizing of the information is readily possible; it is necessary only to associate an appropriate identifying code word with each frequency channel. This is a particularly useful alternative read-out if the data are to be recorded and machine processed.

The unique circuitry associated with this receiver type resides in the individual passive filters defining each frequency subband. Multielement filters are ordinarily employed. Physically, the elements may involve coaxial, waveguide, or strip-line techniques. The complexity of the filter is controlled largely by the filter characteristics desired—the shape of the passband, the off-frequency rejection, the allowable insertion loss, etc.

There are two principal problems. First, a suitable low-loss coupling to the many filter inputs must be made from a single antenna. Second, the design must recognize an ambiguity problem that can arise when signals are received at frequencies near crossover points of adjacent filters. This latter problem is revealed in Figure 9-20. Presuming something less than infinite skirt selectivity (and a nonrectangular passband shape) it is apparent that a strong signal in the filter crossover region can register in more

*The "spanotron" is a microwave tube providing inherently the properties of frequency selection and detection equivalent to the passive filter-crystal detector combination. The latter combination appears to have practical advantages arguing against the use of the spanotron except in special cases. A tunable, single-channel equivalent of the spanotron is furnished by the cyclotron resonance rf detector tube. A comparison of the two (and a complete reference list) is given in Reference 15.
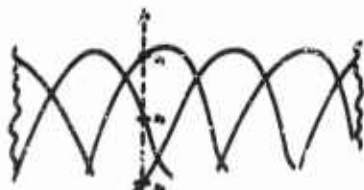
FIGURE 9-20 Possible ambiguity in multiple-channel receiver. A signal at f_c can register in more than one channel (but at different amplitude levels: z_1, z_2, z_3).

than one channel—hence, the ambiguity. Fortunately, it can be expected that the signal levels in the several affected channels will be different and this furnishes the basis for the operation of video logic circuitry which "assigns" a signal to an individual channel indicator by noting relative responses and combinations of responses. Such circuits are sometimes called "ambels" for "ambiguity eliminators".

Since the logic circuit operation is based on relative amplitude, very definite requirements are automatically imposed on standardized gain and response characteristics in video amplifiers, on uniform characteristics in direct-detector elements, etc. This need has fostered the development of video amplifiers of remarkable uniformity over wide ranges of operation—logarithmic video amplifiers being no exception.

This receiver type is the basis for a very useful class of intercep. "systems". A receiver of the form described provides a quantized measure of intercepted signal frequency. Associated with it is a second receiver (Figure
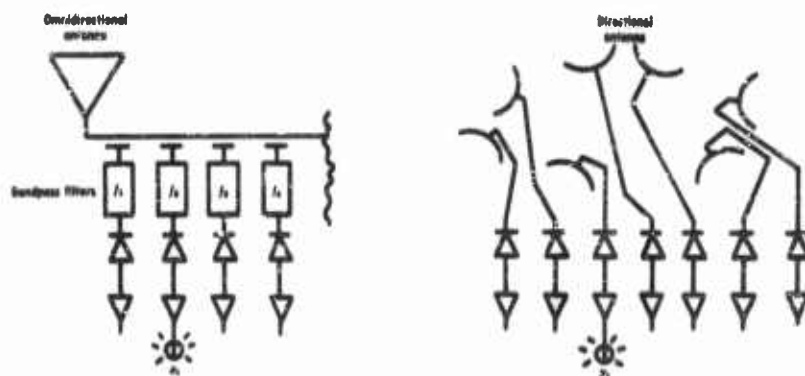
FIGURE 9-21 A receiver for frequency and direction. Time-coincident outputs x_1 and y_1 define frequency and direction of arrival of a pulse.

9-21) wide-open in frequency but quantized in direction; i.e., the several "channels" are each fed by an individual antenna element having directional properties. The several antenna elements provide contiguous coverage over the desired sector (as much as 360 degrees). Thus, an individual signal will register in two channels, one identifying signal frequency, the other direction of arrival. The two responses can be related on a time-coincidence basis to give essentially instantaneous indication of frequency and direction. Further extensions to include some analysis of signal details—prf, pulsewidth, polarization, etc.—have been developed in still more complex devices.

These receivers are particularly adapted to "automatic" operation. An operator is not basically necessary, and the signal indications are directly adaptable to recording (through taping or printing). But these features are obtained only with reasonably high signal-to-noise ratios if an excessive number of false indications is to be avoided; basic receiver sensitivity then is apt to be low—to fall in the -35 to -45 dbm range for S- and X-band units. Individual channel widths range from a few megacycles to several tens of megacycles. While very substantial sensitivity improvements can be (and are) brought about by the inclusion of a broadband r-f preamplifier ahead of the frequency filters, a comparable sensitivity improvement in the "direction" channels would require a multiplicity of such amplifiers (and they would necessarily have to be balanced in gain-frequency characteristics). With the present state-of-the-art this is considered an excessive price; however, some sensitivity improvement is brought about in the direction channels by the higher gain of the directional antennas.

9.8 Superheterodyne Receivers

The history of broadcast and communications receivers shows several receiver forms—crystal-detector, regenerative, TRF—to have preceded the superheterodynes. But the superheterodynes, by virtue of better sensitivity and selectivity, later almost completely displaced the earlier types. There has been some repetition of this trend in the microwave intercept receiver field. Each new band extending the total frequency range generally has been served first by the less-sophisticated receiver types. With the development of the necessary components, the superheterodynes have then followed, again often bringing improved performance in terms of greater sensitivity and selectivity. Although it is of great importance, the superheterodyne, however, has not completely displaced the other intercept receiver types in the microwave ranges nor does it appear that it will do so; high sensitivity and selectivity alone are not the principal requirements for many intercept receiver tasks. In fact, many of the superheterodyne design variations to be described herein are prompted by the necessity for modifying the basic

characteristics associated with that receiver type so as to produce a greater versatility for microwave intercept receiver use.

9.8.1 Low-Frequency Superheterodynes

At lower frequencies—the VHF range and below—the superheterodyne is substantially without competition; its excellent sensitivity and selectivity characteristics are usually of maximum importance. Fortunately, there is an extensive technical background to support the design of these lower frequency receivers. It results from the long period of research and development undertaken to satisfy the requirements imposed on receivers used in the communications and entertainment fields. Most of this technical information is directly adaptable to use in low-frequency intercept receivers.

This opportunity for the utilization of "conventional" techniques comes about largely because of the particular nature of the low-frequency intercept operation. At high frequencies, mere knowledge of the existence of a signal at some particular time or of its location can be tremendously important. Identification of signal *type*, a potentially rapid accomplishment, usually rounds out the intercept operation. But at low frequencies it is necessary to anticipate the continual existence of many signals, and there may well be little about the individual signal type—AM, FM, code, etc.—that immediately identifies it. Determinations of direction of arrival, propagation mechanism, polarization, etc., or of message content (a time consuming process) may be primary system functions as required for signal identification or for intelligence purposes. Thus, it is often necessary to accomplish a fast, cursory examination of many signals in a screening process (and fast-tuning techniques, panoramic presentations, etc., and application). But the follow-up may well be a detailed, long-time examination of a particular signal. This leads to stringent receiver requirements in stability, selectivity, signal-to-noise ratio, etc., that are most easily satisfied by the superheterodyne.

The low-frequency superheterodyne intercept receiver, then, evolves as the natural development of the communications superheterodyne receiver augmented frequently with circuitry providing fast-sweep and panoramic presentation characteristics. It is interesting to note that even these specialized features are not unknown in receivers designed for more conventional uses in the same general frequency ranges.

9.8.2 High-Frequency Superheterodynes

The superheterodyne receivers for higher frequencies—particularly in the UHF and microwave ranges—do not have the broad technical support of techniques developed for other receivers. Admittedly, the receivers designed for incorporation in certain radar systems and in point-to-point communi-

cations links are similar in certain important respects. But the intercept receiver requirements in terms ... broadband coverage, flexible tunability, wide or variable acceptance bandwidth, etc., have been substantially unique and the related circuitry bears description here. It should be noted that trends in modern electronic weapons systems are producing comparable requirements for flexibility in the receivers associated with the newer target location and weapons guidance systems.

The discussion will be divided on the basis of tuning characteristics and acceptance bandwidth as was done with the direct-detection receivers. The order of presentation will be different; the mechanically tuned, sensitive, narrow acceptance bandwidth receiver will be first described, the associated features being easily generated by superheterodyne techniques (in contrast to the related problems encountered in the direct-detection receiver).

9.3.3 Mechanically Tuned Superheterodyne Receivers (Slow-Scan) *

This receiver type results from the natural extension of basic superheterodyne techniques to the higher frequency ranges. The block diagram

FIGURE 9-22 Basic superheterodyne configuration.

is conventional (Figure 9-22). It is a tremendously important intercept receiver type. Though the necessary receiver techniques are basically familiar, the extension in frequency and operating flexibility is not always easily accomplished. In particular there are often new design demands on pre-selectors, local oscillators, mixers, and i-f amplifiers.

This receiver type admits to a wide selection of local oscillators; triodes, carcinotrons, backward-wave oscillators, klystrons, etc., all find use. Though it is not fundamentally necessary, the manually tuned receivers often feature narrow acceptance bandwidths (of a few megacycles) and this characteristic usually brings an accompanying requirement for local oscillator stability if the receiver is to be operated at all on a fixed-tuned basis. Fortunately, this stability requirement is not inconsistent with the features of mechanically tuned local oscillators. It also is generally true that the allowable tuning rates, in megacycles per second per second, are then apt to be low. These receivers are frequently fitted with a motor-drive arrangement permitting a repetitive scan of the total frequency range, or of a limited sector. Again, scan rates are usually low—a few cycles per second at the maximum if the

*The classic example is the AN/APR-9. The modern derivation is described in Reference 16.

primary local oscillator is involved in the tuning process. An excellent example of this technique is covered in Reference 17.

Double- and triple-detection techniques find application in these "conventional" superheterodyne receivers. Of particular interest is the development of the panoramic display via a second, electronically tuned local oscillator which furnishes a rapid "scan" of the signals in a wide, first i-f amplifier (Section 34-7, pages 984-987 of Reference 2).

The tuning characteristics of the receiver are largely set by the tuning versatility of the local oscillator. At some risk it can be generalized that the oscillators inherently more difficult to tune (in the technical sense) provide advantages in stable operation. The more versatile the receiver tuning, as via the electronically tuned oscillators, the lower the stability (a disadvantage offset in many instances by the greater tuning flexibility). Local oscillator r-f output power requirements are conventional—a few milliwatts, typically, except in special instances. More total output power is required in certain receivers utilizing multiple local oscillator signals. These signals are sometimes developed by modulating a higher-power local oscillator so as to produce an appropriate spectral distribution of sidebands, each serving as a "local oscillator" signal.

No unusual mixer techniques are introduced necessarily by the intercept receiver implication. The i-f amplifier center frequencies and bandwidths are often typical of microwave receivers in general. Mixer conversion losses in the order of 6 to 10 decibels at the S-band, and 8 to 12 decibels at the X-band are usual with i-f amplifier center frequencies in the 30 to 200 megacycles per second range. However, image rejection problems tend to be accentuated and this has a reaction on mixer design in that unusual selections of i-f amplifier center frequencies are suggested. The spurious response problem is discussed with particular reference to intercept receivers in Section 34-8, pages 987-992 of Reference 2. Thus, center frequencies of several hundred or several thousand megacycles are sometimes employed. (This aids the frequency separation of responses and assists the selection of the desired response by a preselector of practical selectivity characteristics.) The reaction on mixer design is serious usually only if the ratio of input signal frequency value to the translated intermediate frequency counterpart is low.

As an alternative to the high i-f amplifier frequency, the center frequency is sometimes lowered to bring the two responses together in frequency (the zero-frequency superheterodyne). No preselection is employed on the presumption that, since the impaired resolution can be tolerated, the closeness in frequency of the responses will aid their identification. This simplifies mixer design in one sense (the i-f amplifier frequency is low). However,

the system now must contend with possible increased (low-frequency) noise contributions from the mixer and local oscillator. It complicates it in another sense, for if the i-f amplifier response extends to low frequencies, the amplifier will pass signals developed through the direct-detection action of the mixer crystal (in the manner of the crystal-video receiver). Though the efficiency of the direct-detection process is low, a substantial response can be developed by a strong signal. Since the direct-detection action is not conditioned by any frequency selection process, a very substantial confusion can result—a direct-detected response may be erroneously related to the current frequency setting of the receiver which is correctly applicable only to signals produced by the normal superheterodyne conversion and detection process. One common remedy affects mixer design: by the use of a balanced mixer the output signal resulting from direct-detection action can be very substantially reduced. Alternatively, a direct-detection comparison channel can be utilized. A third approach involves r-f compression of dynamic range before mixing so as to hold direct-detection spurious re   ses below the noise level.

Much of the attraction of the superheterodyne intercept receiver resides in the characteristics set by the i-f amplifier design. In particular, the pass-band characteristics and off-frequency signal rejection abilities are of paramount importance. This has led to a great deal of attention to i-f amplifier design intended to accentuate features useful in intercept receivers—low-noise characteristics, very high or very low center frequencies, logarithmic response characteristics in some cases, wide dynamic range, etc. When the requirements are inconsistent for a single amplifier, the usual double- or triple-detection principle often affords the answer. The i-f amplifier requirements for the manually tuned narrow acceptance bandwidth receiver are largely conventional. The more important i-f amplifier design variations are described in Section 9.8.4 in connection with electronically tuned super-heterodynes.

One of the problems accentuated in microwave intercept receivers is that of preselector design. The receiver that is manually tuned (by hand or by motor-drive) is favored in this regard in that there is then admitted for consideration the multiple-circuit, mechanically tuned preselector. In combination with high-center-frequency (VHF) i-f amplifiers (to separate responses), multiple-circuit preselectors can bring about image and spurious response rejections in the 60 to 80-decibel range. It is desirable that the insertion loss of such a passive preselector be low; values of 0.5 to 3 decibels are usual (Chapter 28, pages 741-795 of Reference 2 and see also Reference 18).

There are no special considerations attendant to the development of the

second detector and video amplifier portions of the superheterodyne intercept receiver. A signal once established in the i-f amplifier can be treated by conventional techniques. It is necessary that the response characteristics (dynamic range, transient response, etc.) match the requirement of the receiver. But many of the problems in this regard (in developing high gain, tailored response characteristics, etc.) can be handled more conveniently in the i-f amplifier of the superheterodyne than in the postdetection circuitry of the direct-detection receiver. The unusual problems that do exist are largely generated by the possible multiple-purpose uses for the receiver—perhaps the need to handle in quick succession narrow pulse signals, c-w signals, FM signals, etc. As a consequence, a single receiver may involve a conventional second-detector for AM signals, an FM discriminator, SSB circuitry, a wideband video amplifier to drive a pulse analyzer or video recorder, an audio channel for headphone or tape use, digital read-out circuitry, etc. These abilities may all be in use simultaneously or individually by option. There is a resultant circuit complication but no unusually difficult design problem uniquely ascribable to the intercept function.

9.8.4 The Electronically Tuned Superheterodyne

Versatility in tuning is the keynote of the electronically tuned superheterodyne. Basic to this receiver type is an electronically tuned local oscillator—a carcinotron, backward-wave oscillator, helitron, ophitron, voltage-tuned magnetron, etc., in the microwave range; usually a voltage-sensitive capacitor or current-sensitive inductor in the frequency-determining circuit of a more conventional local oscillator in the lower frequency receivers. It is this electronically tuned element which distinguishes the receiver type; with this substitution, the block diagram of Figure 9-22 is applicable.

With this tuning versatility come certain advantages for some intercept receiver uses. The practical limitation on scan rate largely vanishes; the frequency scanning process can be programmed easily over either full or selected incremental r-f bandwidths; the frequency scan can be stopped quickly and perhaps automatically on a signal-triggered basis; there is a convenient voltage or current versus frequency relationship aiding automatic frequency read-out, etc. (Reference 19).

There are problems introduced or accentuated by the electronic tuning process. A high scan rate argues for receiver acceptance bandwidths that push the bandwidth capabilities of conventional i-f amplifiers; there are not now readily available the electronically tuned preselectors having all of the characteristics in bandwidth, skirt selectivity, noise figure, etc., ordinarily desired. (The backward-wave amplifier offers great promise in this direction.) These factors push i-f amplifier design to very low or very high center frequencies (for control of spurious responses), and to substantial acceptance

bandwidths (for higher intercept probability in some typical search conditions).

Microwave i-f amplifiers are sometimes used. There is then no problem in developing wide bandwidths (and bandwidth adjustment is readily accomplished through the insertion of passive, bandwidth limiting filters).* By an appropriate design of the intermediate frequency amplifier—for example, by selecting a microwave center frequency value greater than the width, per band, of the radio frequency ranges to be monitored—problems from images and spurious responses can be largely eliminated despite the absence of a preselector. It should be noted that some very ingenious schemes have been employed successfully in receivers to eliminate an image response without preselection by, in effect, characterizing the multiple responses such that the one desired can be selected by subsequent circuitry. One device accomplishes this by a combination of balanced mixing and broadband phase shifting which establishes a particular signal polarity (assisting selection) for the desired response (Reference 22). Further, it is convenient to cover two r-f bands per local-oscillator tuning range via the simultaneous or alternate use of both higher-signal-frequency and lower-signal-frequency responses, both of which appear within the given i-f amplifier range. A block diagram of such a receiver appears in Figure 9-23. It is possible, also, to multiplex several narrower heterodyned r-f channels through a single, very wide band i-f amplifier in certain cases. This technique was investigated extensively in connection with the development of the ALR-7 receiver (Reference 21).
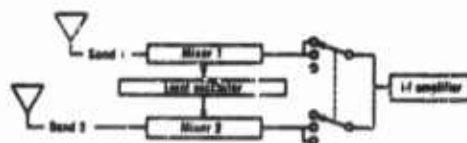
FIGURE 9-23 A single local oscillator—two-band superheterodyne. Example: Band 1 covers 9-15 kc, Band 2 covers 15-21 kc; local oscillator covers 12 to 18 kc; i-f amplifier center frequency set at 3 kc.

The use of a microwave i-f amplifier usually introduces a penalty in terms of noise figure.† This is not necessarily true if a parametric amplifier (with its narrow acceptance bandwidth) is involved. More typically, the amplifier performance is based on the poorer noise characteristics of a much wider band, low-noise TWT. In fact, several tubes may be employed in cascade if high-gain, wide dynamic-range characteristics are required. Fortunately, too, the TWT amplifier is subject to versatile gain control measures which aid in the handling of signals of widely different amplitudes.

The principal reaction on mixer design of the electronic tuning process

*Some of the over-all features are outlined in Reference 20 and a very versatile form of this receiver is described in Reference 21.

†Unless r-f preamplification is available (see Section 9.8.5).

results when a microwave i-f amplifier is to be incorporated in the receiver. This calls for a wideband heterodyned signal bandwidth at a center frequency that may be close to the r-f signal frequency band; both factors, bandwidth and the closeness in frequency, complicate mixer design. An interesting problem can arise if the frequency scan rate, in megacycles per second, is increased to very high values (as in the "microsweep" receiver described in Section 9.9). The variable (with frequency) noise contribution from the mixer then can demodulate in the postdetection circuitry as an interfering audio or video "noise" signal, and its presence complicates true-signal recognition and identification processes.

The electronic tuning process, as such, does not introduce additional factors conditioning the design of the second-detector and postdetection circuitry, except insofar as wide bandwidths may be required (Reference 23).

9.8.5 The Use of R-F Preamplification

As in the case of the direct-detection receivers, it is possible to precede the superheterodyne mixer or passive preselector with r-f preamplification. A tunable preamplifier can substitute for both a passive preselector and a broadband preamplifier. The benefits and problems are entirely comparable, and the over-all receiver performances (TRF versus superheterodyne), for a given acceptance bandwidth, are apt to be very similar.

In some cases, there are important side benefits to the superheterodyne brought about by the use of a preamplifier. Mixer and i-f amplifier design problems which limited receiver sensitivities are removed since the noise figure can be set by the preamplifier. There is burnout protection for the mixer crystal (because of preamplifier strong-signal limiting), and there is some additional reduction in local oscillator radiation (because of the unidirectional preamplifier transmission). The advantages of the superheterodyne configuration in terms of selectivity are retained. In turn, there are introduced the usual potential preamplifier problems discussed in Section 9.7.4 in connection with the direct-detection receivers—the generation of spurious signals in an overdriven preamplifier, the additional system complication, a current lack of entirely suitable electronically tuned preamplifiers, etc.

9.9 The Microsweep Superheterodyne

An extreme extension of the electronic tuning technique exists in the so-called "microsweep" superheterodyne. These receivers take advantage of the fast sweep rates—thousands of megacycles per second per *microsecond*—that are possible with some types of microwave oscillators. (Such tubes are described in Chapter 26). Otherwise, the basic form of the receiver is reasonably conventional.

9.9.1 The Basic Receiver*

With submicrosecond sweep times, it becomes possible to scan the receiver in frequency at a rate such that the full radio-frequency range assigned to the receiver is covered in the time duration of typical radar pulses. The intercept philosophy is discussed in Section 9.5 as the case where $t_s < t_w$. In short, it would be expected that some energy from each radar pulse within the range covered—regardless of pulse frequency—would be intercepted, and that a high pulse-by-pulse intercept probability would result. The high sweep rate may require even wider bandwidths than would be required on the basis of the desired resolution capability or pulse fidelity of the receiver, so that some sacrifice in sensitivity and frequency resolving capability is to be expected.

By establishing a standard r-f input signal level (as through the use of a broadband, limiting preamplifier), a remarkable resolution can be developed —0.5 percent over a 2:1 range at S-band, for example. The limiting process sometimes produces spurious responses when the limiter is confronted with simultaneous, strong pulses on several radio frequencies (a situation produced by several types of multiple-frequency radars).

9.9.2 The Compatible Filter Superheterodyne

The problems associated with the basic microsweep receiver as described have led to the development of somewhat related forms of receivers still involving very rapid sweeping techniques and high intercept probability for many signal types but with important divergences in concept. This basic technique (for a German-made early suggestion of this idea see Reference 25; see also Reference 26) involves first mixing the r-f signal with a linearly sweeping local oscillator; the resultant i-f signal varies linearly in frequency with time. This signal is then fed to a dispersive filter whose transit time is a function of frequency—the frequency applied first (as the sweeping frequency enters the filter passband) must experience the greatest delay. By matching the filter delay characteristics to the sweep rate, the i-f component energies at the output tend to combine at a particular time (hence the term "time-compressing," which is often used to describe the filter properties). The time at which this build-up occurs at the output is a function of the input frequency, thus allowing a measurement of r-f frequency via a time measurement (often with respect to the start of the local oscillator sweep).

The i-f signal in this type of receiver is sweeping in frequency. Unlike most received signal properties, this frequency change is under the control of the receiver designer (through control of the local-oscillator frequency sweep). By designing an i-f filter having characteristics appropriate to this

*Early work in this field is described in Reference 24.

frequency change ("compatible"), the unique properties of the receiver are realized.

The dispersive filter can be realized in either a continuous (Reference 27) or a discretely segmented form.* In one discretely segmented type, a number of conventional filters is used in conjunction with appropriate delay lines as shown in Figure 9-24. The several passbands are selected so that in
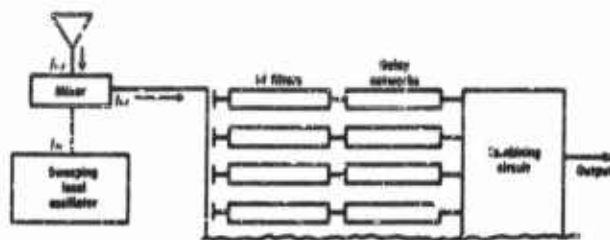
FIGURE 9-24   Compatible filter superheterodyne.

parallel combination, a contiguous frequency coverage is provided across a total i-f bandwidth much wider than the incremental bandwidth of any one filter channel alone. The sweeping action of the single local-oscillator signal f_{lo} is such that the changing difference frequency $f_{i,f}$ produced by the local-oscillator energy in combination with a given fixed frequency signal $f_{r,f}$ is transmitted successively by the several filters.

The filters, in turn, feed time-delay circuits having tailored characteristics (determined by sweep rate) such that all delayed outputs arrive simultaneously at a combining point. The output signal of the channel first excited is delayed the longest, etc. The time of the peak of a response with respect to the initiation of the frequency sweep is a measure of the radio frequency of the signal producing the response. Depending on the combining method (addition, multiplication, etc.) certain performance advantages can be emphasized. A total resolution substantially better than the bandwidth of a single i-f amplifier channel can be produced. In the case of envelope multiplication, the combined output signal (in terms of resolution) of n channels is comparable to that generated by an amplifier of n synchronously tuned stages each having the bandwidth of a single channel. More generally, the limit of resolution is that determined by the impulse response (time buildup) appropriate to the total input signal energy bandwidth accepted by the several filters collectively, which in turn is related to a frequency resolution in terms of the sweep rate of the local oscillator.

*A discussion of various realizations of both forms of filter may be found in W. D. White, patent #2,882,395 dated April 14, 1959.

If the signal time-duration is sufficient to allow the successive excitation of responses in all channels, the subsequent addition of signal outputs is made on an integrating basis more favorable than that controlling the combining of the random noise powers of the several channels; the result is a signal-to-noise ratio enhancement. (The resulting sensitivity can be as good as that developed in a fixed-tuned receiver of comparable resolution and noise figure developed through normal design practices.) Finally, some discrimination against image and spurious responses results from this circuit form since only the combinations from a real signal add correctly (in time) at the combining point to produce a recognizable response; relative channel delay times are incorrect for the additions of most spurious signals.

9.9.3 The Ratio-Sweep Receiver

An interesting member of the fast-sweeping receiver group is described as the ratio-sweep receiver (Reference 28). In contrast to the one local oscillator, multiple-amplifier-channel configuration of the "compatible filter" receiver, this device employs multiple local oscillators but with only one i-f amplifier channel per local oscillator. All such amplifiers are identical as to bandpass characteristics.

In simplified form, two local oscillators, triggered from a common starting point, are swept at different rates across the band to be monitored (Figure 9-25). The responses developed by a given fixed-frequency signal are then generated at different times in the separate i-f amplifier channels with the time difference ($t_2 - t_1$ in Figure 9-25) being a measure of signal frequency.

A particularly useful elaboration makes use of making matrixes of mixed-base congruent numbers to solve some important practical problems that arise in assuring a useful frequency resolving power. Each local oscillator
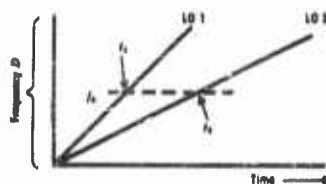
FIGURE 9-25 Simple "ratio sweep" receiver. A signal (f_o) produces a response in the two i-f channels associated with the two local oscillators. The time difference $t_2 - t_1$, is related to f_o.
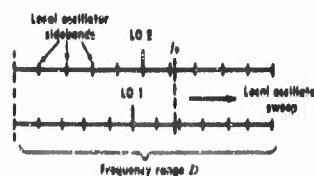
FIGURE 9-26 Ratio sweep receiver employing two sets of local oscillator signals with the total band, D, covered by 10 and 11 components, respectively. The time difference of the response uniquely measures f_o. The oscillators sweep a range $D/10$ and $D/11$, respectively.

signal is replaced by a set of uniformly spaced local oscillator signals swept in coincidence. A group is produced by modulating a given high-level oscillator signal so as to generate, via sidebands, the set of equally spaced components in a "comb" arrangement. The number of components used defines the number of elemental "states" in the system. A single i-f amplifier is associated with each group.

In practice, several groups of local oscillator signals are swept simultaneously (and rapidly) through the frequency limits necessary to monitor a given r-f range. For example, two groups might be used (Figure 9-26); there would then be a total of two associated "i-f" amplifiers. The number (base) of local oscillator signals per group would differ. Their frequency spacings and sweep limits, too, would differ such that the total assigned r-f band, D, would be covered by each group. Thus, if the number of local oscillator signals per group were m and n, their spacings would be D/m and D/n respectively (which would also define sweep limits) and sweep rates would be adjusted for equal time coverage by the groups of the total band D. Beyond this, there is an optimum selection of m and n to optimize the frequency resolution. The total number of elemental states is the sum of the bases ($m + n$) while the resolution is related to the product of the bases ($m \times n$).

Note that because the several spaced local oscillator signals are swept in a group, the total r-f range covered is m (or n) times the local oscillator frequency spacing (if there are m or n such signals), i.e., the necessary extent of the frequency sweep per local oscillator is greatly reduced for a given r-f range. For example, two groups might be used (Figure 9-26); given total frequency coverage if the number of local oscillator signals per group is moderately large. This has important benefits in allowing favorable reductions in primary local oscillator sweep time and in sweep voltage amplitudes without adversely affecting the inspection time (dwell time) and resolution for a given signal.

9.10 Miscellaneous Receivers

9.10.1 Instantaneous Frequency Indicators

The non-scanning, multiple-channel receivers described in Section 9.7.5 are a form of instantaneous frequency indicator. They are always sensitive to signals anywhere in the band to be monitored; they have selectivity and resolution to a degree set by the individual channel characteristics. The resolution is on a quantized basis and there is no additional direct resolution within a channel width (as in the case of the fast-sweeping receivers of Sections 9.9.2 and 9.9.3). There are a number of receiver types which seek

the virtues in intercept probability of continuous (non-scan) monitoring but with a continuous (rather than quantized) resolution in frequency developed with the aid of a dispersive circuit element. The term "non-scan spectrum analyzer" is sometimes applied to these devices.

All employ some frequency-sensitive element (other than the classic fixed-frequency tuned circuit) which generates a frequency-related variation in output amplitude or phase or transit time. For example, the frequency sensitive standing wave pattern established in a deliberately mismatched r-f transmission line (or waveguide) can be sampled so as to determine frequency via the calibrated ratio of pattern amplitudes at the fixed sampling points (Reference 29).

In another form, the incoming r-f signal is divided between two transmission channels, one containing a frequency-sensitive attenuator, the second providing a constant loss across the band (Reference 30). Again, the ratio of the detected amplitudes of the signals in the two channels can be used as a measure of frequency (Figure 9-27).

There are important problems common to these devices. The frequency determination involves an amplitude measurement so accuracy tends to be sensitive to amplitude variations in incoming signal level; however, broadband, amplitude-limiting r-f preamplifiers (available through TWT techniques) offer an important assistance. The data are conveniently read out on visual displays but the direct output data ordinarily are not in an optimum form for automatic machine handling, some additional processing is required. The systems usually are in difficulty if more than one signal frequency is present simultaneously. Measurement accuracy is usually dependent on maintenance of "balance" in characteristics of two or more channels, r-f, or video. This imposes severe problems on circuits and components in the face of variations in signal levels, temperature, component aging, etc.

A novel answer to the dynamic range and channel amplitude balance problems lies in the use of a transmission element having a time delay which is frequency sensitive (a waveguide operated in the near cutoff region, for example). If a transmission time comparison is made of samples of a signal sent through channels having transmis time characteristics respectively constant with frequency and frequency sensitive, a measure of frequency can be obtained. The time comparison can be based on an r-f phase measurement. However, one particularly interesting form of receiver measures the envelope delay of a high frequency (perhaps 50 megacycles) modulation superimposed on the r-f signal (Reference 31).

As a general rule, all of these receivers trade a significant loss in sensitivity for the unique frequency resolving and continuous monitoring properties. Broadband r-f preamplifiers offer an important remedy (except for the pen-
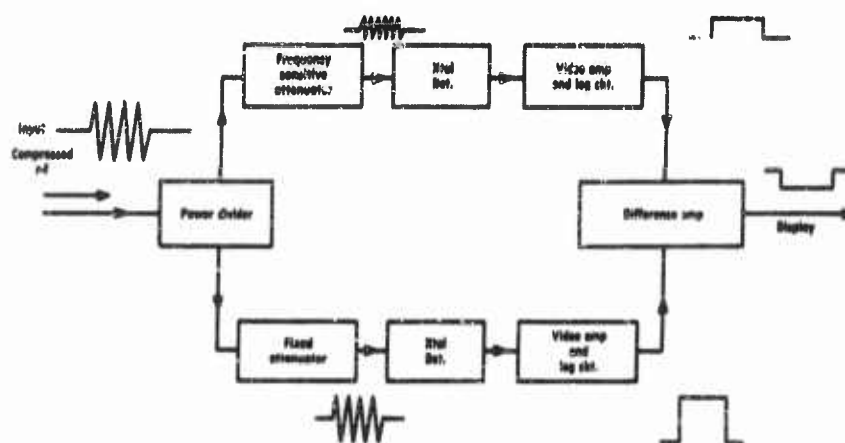
FIGURE 9-27   An "instantaneous" frequency discriminator.

alty to system simplicity), and their use results in practical sensitivities consistent with the preamplifier noise figures and effective bandwidths of the total system.

9.10.2 Superregenerative Receivers

The superregenerative principle has been applied with moderate success in specialized microwave receiver applications (pages 545-579 of Reference 1). The possibility of developing very large r-f gain in a regenerative stage is always attractive—particularly at microwave frequencies. Unfortunately, the generally poor stability and the critical adjustment identified with superregenerative circuits work against any general usage; problems are emphasized if the receiver must be easily tuned. The advent of new components offers promise of better performance (References 32 and 33).

There appears to be an interesting involvement of the superregenerative principle with the parametric preamplifier for receiver use. There can be an apparent enhancement of acceptance bandwidth brought about by a com-

bined AM-FM (sweeping) action resulting from a high "quench" frequency. There has been reported (Reference 34) a superregenerative parametric L-band amplifier with a "quench" frequency of 0.25 megacycle, an r-f acceptance bandwidth of 2 megacycles, and a noise figure of 1 decibel. This unit had a stable gain of 56 decibels and an equivalent video (or information) bandwidth on the order of 0.25 megacycle. By taking full advantage of integration, the pulsed signal sensitivity was better than -114 dbm. Stable gains as high as 80 decibels have been attained, with a corresponding reduction in bandwidths.

9.11 Special Components and Circuits

Countermeasures receiver designs furnish interesting examples of the dependency for advances in intercept receiver technology on new components and new circuits. In the recent past, the advents of TWTs and distributed r-f amplifiers have been reflected in new receiver abilities. So, too, have the developments in local oscillators—the voltage-tuned magnetron, carcinotrons, backward-wave oscillators, etc., have found immediate applications in the microwave ranges. There have been comparable influences at lower frequencies resulting from the introduction of voltage-and-current-sensitive capacitors and inductors. Ferrites have brought new abilities in isolators, modulators, choppers, etc. There are new forms of antennas, transmission systems, waveguides, solid-state components, etc.

A major profit can be anticipated from the utilization of such new devices, but there are usually both advantages and disadvantages which tend to limit broad usage. Thus, the maser has a valuable application in certain fields; but the limited r-f bandwidths and the current practical complications in usage argue against a general intercept receiver utility—particularly when the extreme low-noise properties may be lost in the presence of other important system noise sources (Reference 35). The parametric amplifier offers great promise; it still has strong competition from the low-noise TWT when the latter's wide bandwidth, nonreciprocal transmission characteristics, and ease of utilization are considered. Nor does the parametric amplifier currently enjoy an important noise-figure advantage in the high microwave range. Its basic simplicity is partially offset by the need for a "pump"—currently a vacuum tube. So each new device must be carefully evaluated. Each can be anticipated to affect countermeasures receiver design to a degree, and it is important in reviewing countermeasure receivers to consider, as well, the current information on new component developments.

While the receiver design dependency on new components is obvious,

there is an important reaction in the reverse direction. Certain important developments in circuits and components, in turn, are traceable in part to work initiated by the needs of countermeasures receivers. These needs, for example, account for a particular attention to certain broadband, low-noise characteristics in i-f amplifiers, and to unusual i-f amplifier response and gain characteristics. The same is true in video amplifiers; direct-detection receivers in particular impose stringent requirements on low-noise features, logarithmic characteristics, "square-root" response, etc. Both slow and instantaneous agc circuitry is often important. The success of several receiver techniques is dependent on the availability of effective, broadband r-f limiters. Choppers and pulse stretchers have required particular attention; and the peculiar needs in direct detectors and in mixers have generated related research and development utilizing a wide variety of nonlinear devices. This interrelationship is emphasized by the frequent references in this chapter to the material on special circuits and components, Chapters 24 through 28.

9.12 Some Miscellaneous Factors Affecting Intercept Receiver Designs

This chapter has been largely concerned with receiver design problems generated by the unique natures of some of the tasks of an intercept receiver. The resultant requirements imposed on sensitivity, selectivity, dynamic range, etc., have been emphasized. There are additional operational factors, no less important on some occasions, that require the introduction of unusual receiver features or of unusual methods of operational employment. While there are many examples, only a few representative types will be mentioned here. These problems are not unique to any one receiver type. However, some receiver types are better equipped, fundamentally, to deal with some of the problems.

A factor which plagues the optimum utilization of an intercept receiver is the lack of knowledge of the precise nature of the signal to be intercepted. Optimization of bandwidths and sweep rates, the utilization of integration techniques, etc., are largely ruled out unless a particular signal of known characteristics is to be sought. It is more likely that the receiver must be prepared to handle a variety of signal types. The "variety" may be represented by much more than small variations in the "standard" parameters (pulsewidth, prf, etc.) of a given emitter; it may well include large changes in r-f, prf, polarization, etc., as well as contrasting modulation types—c-w signals, FM signals, coded pulse groups, SSB, etc. This has led to the development of subcircuits for alternate functions—choices in detectors, amplifiers, etc.—to be utilized simultaneously in parallel or to be inserted into a

receiver on a module exchange basis in accordance with some current operational intent. It is sometimes possible to equip a receiver with only a moderate additional circuit complication which purchases a substantially improved ability to handle simultaneously some sharply contrasting signal forms.

9.12.1 C-W Reception

Simultaneous reception of pulse and c-w signals provides an example of the above concept. A conventional pulse receiver can be provided with an enhanced sensitivity to c-w signals through the addition of a conventional "chopper" (to modulate the c-w signals with the chopper frequency) plus a narrow-band postdetection amplifier centered at the chopper frequency. Note that it is highly desirable that the chopper be introduced into the receiver circuitry at a point such that the c-w signals are modulated (chopped) but the predominant noise is not. Otherwise, both signals and noise will be characterized by the chopping action and the ability of the narrow, postdetection amplifier to discriminate against the noise will be seriously penalized. This circumstance is not unique to intercept receivers. The enhancement of c-w reception comes through the reduction in noise bought by the bandwidth constriction in the narrow-band postdetection "c-w channel"; the c-w signal-to-noise ratio is consequently improved. The c-w channel width can be very narrow—a few cycles, perhaps. The signal-to-noise ratio improves with narrower values; the response time is lengthened thereby (the data rate is reduced). An important limit to the channel narrowing is set in scanning receivers since the tuning action causes short "samples" of the chopped c-w signal to be delivered to the postdetection circuitry. This is discussed below. There is a small penalty to pulse signal reception resulting from the periodic loss of signal from the chopping action; for that reason, the c-w reception feature is usually made optional.

It is interesting to note that if the receiver is operated on a constant repetitive frequency-scan basis, a single postdetection "narrow channel" bandwidth value (β') can be selected to optimize c-w reception in accordance with the "on-frequency" time increment as set by scan-rate and the *predetection* bandwidth α. That time interval fixes the duration in the postdetection circuitry of the c-w signal response per receiver scan. In a receiver intended for both pulse and c-w signal detection, the predetection bandwidth selection α (for example, the i-f amplifier acceptance bandwidth) can be based purely on the needs for pulse reception and a second, standard postdetection amplifier (of bandwidth β) for pulse signals can be paralleled with the narrow (β') amplifier. The available receiver bandwidth combinations (using the single predetection bandwidth) can then be made optimum for

both c-w and pulse signals. The predetection bandwidth could be narrowed in the expectation of increasing signal-to-noise ratio for c-w signals; but this "advantage" would be offset in a scanning receiver by a necessary widening of the c-w channel postdetection bandwidth (β') as required to accommodate the now shorter (in time) c-w signals delivered to the postdetection circuitry. This is in accordance with the expression for effective noise bandwidth, Eq (9-10).

9.12.2 Intercept of "Variable-Frequency" Radars

The manner in which the radio frequency of an intercepted signal is handled is a matter of major importance in the design and evaluation of an intercept receiver. This importance is based not only on historical precedent, but the evident fact that whatever r-f processes are to be accomplished in the receiver must be carried out before envelope detection—in the front end of the equipment. Selection of appropriate video processes, in turn, must be compatible with the r-f design of the receiver. This criterion has been used as the primary basis for classifying a wide variety of receiver types in this chapter.

Thus, a fundamental receiver concern arises with the present trend toward radars capable of operating on more than one frequency. In some such radars, changes in transmitted frequency can be made only over narrow limits at rates ranging from seconds to hours. These systems would pose only moderate problems for any of the frequency-scanning receivers described in this chapter. However, radars do exist which can vary their transmitted frequency (on a programmed or random basis) over wide ranges—in some cases during an interpulse interval. Other radars emit several pulses on different frequencies simultaneously or in a short burst of pulses on different frequencies. Such signals cannot usually be handled adequately—in terms of collecting sufficient information for analysis—by any of the scanning receiver types that have been discussed. The main difficulty is the inability to attain a reasonable probability of intercepting a sufficient number of consecutive (or simultaneous) pulses, which occur at different frequencies. A wide-open receiver, on the other hand, could be expected to intercept all of the pulses, although the collected data would provide no clue as to the unusual nature of the transmitter frequency characteristics.

Some question could well be raised as to the significance of measuring the frequency of a variable-frequency emitter, especially in the case of random changes. However, even though accurate frequency measurements were not deemed necessary, there remain certain questions to be answered: Is the frequency being varied? At what rate? Over what limits? Is the variation programmed or random?

Fortunately, appropriate combinations of conventional techniques and special circuits have been devised which do a remarkable job of intercepting and tagging data from many of the existing variable-frequency signals. As other radar types are developed, it will inevitably be necessary to devise new special circuits to handle them.

Despite the trend toward frequency variations in many of the newer radars, the bulk of the signals of interest to intercept receivers are of the single-frequency type—and thus amenable to the techniques described in this chapter.

9.12.3 Other Signal Recognition Problems

Besides the variable-frequency capability, there are many other "unusual" radars whose characteristics require special receiver techniques or circuit combinations for adequate intercept, identification, and detailed analysis. A few examples are cited below. Pulses may be transmitted in coded groups. Pulse repetition frequency may be programmed or randomized over wide limits. Complex scan modulations can be employed. Coding filters can be used to modulate the radar carrier frequency so that the transmitted signal has a "noise-like" appearance during the pulse unless it is fed through a suitable decoding filter. No list of such characteristics would remain complete for very long since new radar techniques are continually under development; combinations of such characteristics are also possible in a given emitter. These increases in radar complexity require corresponding sophistications in intercept receivers that must operate in an environment containing these signals. There is evidence that development of successful receiver techniques can keep pace with the development of exotic characteristics in new radars. The point here is that such tasks exert important influences in intercept receiver design.

9.12.4 Receiver "Look Through"

In some circumstances, an intercept receiver is called upon to operate in the presence of emissions—usually from a friendly radar or jamming signal —that are not of interest in the ELINT collection task but which can degrade the performance of the equipment. The final ability to recognize a desired signal—to look-through—is set by a signal-to-noise ratio, where the noise in this case is not ascribable to the receiver itself, or even to natural causes.

If no control over the transmitter can be exercised, some improvement can often be derived by optimizing the signal-to-noise ratio through utilization of the most favorable (discriminatory) antenna patterns. For noise jamming signals, there is often little else than can be done, since the noise

discrimination techniques decribed in connection with c-w signal reception (Section 9.12.1) are not applicable here. The noise is developed external to the receiver; it arrives as a separate signal, and the modulation (chopping) process characterizes signals and noise alike so that a subsequent selection of the signal is not aided.

In the radar (low duty-cycle) interference case, a special receiver trained particularly on the offending signal can sometimes be used to gate off the collection receiver during the moments of interference. The gating can be on a signal amplitude basis or on a signal coincidence basis.

Occasionally, frequency elimination filters are useful in interference reduction.

There are other techniques that can be employed when there is some opportunity to operate in cooperation with the undesired transmission. For a jamming noise source, a synchronized time-gating process is sometimes used such that the receiver is gated on for short (fractional-second) intervals, perhaps randomly spaced in time and duration, during which the noise source is gated off. The duty-cycle of the receiver (the on time) is low but it usually is sufficient for the short inspections of the signal environment necessary during a jamming operation.*

A somewhat related process can be utilized when the receiver suffers interference from a cooperating radar. The receiver is gated off momentarily at the instant of the radar transmission by a synchronized blanking pre-pulse delivered from the radar to the receiver, perhaps by video cable.† In this case, there is little penalty to receiver duty-cycle.

More sophisticated techniques have been proposed and sometimes employed with moderate success. Fortunately, the antenna discrimination technique is often of major benefit when patterns can be arranged appropriately (as may be possible in a joint jammer-receiver installation in an aircraft, for example). If a noise jamming source is some distance away, there may be relatively little difficulty. In fact, the desired detection of noise jamming signals can present major intercept problems when standard receivers with low receiving antenna gains are employed.

*The duty-cycle must be low since appreciable time gaps in the jamming would reduce jamming effectiveness.

†This is the "interference blanking" technique common to radar operation

REFERENCES

1. S. N. Van Voorhis, *Microwave Receivers*, Radiation Laboratory Series, Vol. 23, McGraw-Hill, New York, 1948.

2. Radio Research Laboratory Staff, *Very High Frequency Techniques*, Vol. II, McGraw-Hill, New York, 1947.

3. *Directory of Intercept and Analysis Equipments*, Section VIII, Report No. 63.6-F, 2nd ed., Contract No. OA-36-039-SC-73225, Haller, Raymond and Brown, Inc. (currently HRB-Singer), 30 April 1958.

4. *Measurements of Radar Radiation Near the Zenith*, Report No. 159-F, HRB-Singer, 17 September 1958.

5. *Signal Density Study, Surface Radar Deployment*, A. L. Hiebert et al., Rand Report No. R-280, 1 September 1955.

6. *Signal Density Study II, Zone of Interior*, A. L. Hiebert et al., Rand Report No R-310, June 1959.

7. J. I. Marcum, *A Statistical Theory of Target Detection by Pulsed Radar*, RAND Corporation, Parts I and II.

8. W. B. Davenport and W. L. Root, *Random Signals and Noise*, McGraw-Hill, New York, 1958.

9. S. Goldman, *Frequency Analysis, Modulation and Noise*, McGraw-Hill, New York, 1948.

10. R. C. Cumming, *Sensitivity Limitations of Microwave Receivers*, Report No. 703-2, Stanford Electronics Laboratories, 1960.

11. J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals*, Vol. 24, Radiation Laboratory Series, McGraw-Hill, New York, 1950.

12. J. L. Grigsby, *An Investigation of the Sensitivity of Some Receivers Having Large RF Bandwidth*, Technical Report No. 553-1, Stanford Electronics Laboratories, 4 November 1958.

13. W. K. Ayer, *Characteristics of Crystal Video Receivers Employing R-F Pre-Amplification*, Technical Report No. 150-3, Stanford Electronics Laboratories, 20 September 1956.

14. *Final Engineering Report on Electronics Reconnaissance Set AN/DLD-2(XA-1)*, Report No. 3642-1, Airborne Instruments Laboratory, Inc., July 1958.

15. F. M. Turner, "A Broad-Band Cyclotron Resonance RF Detector Tube," *Proc. IRE*, Vol. 48, No. 5, May 1960.

16. *Handbook Service Instructions Receiving Set AN/ALR-3*, NAVAER 16-30 ALR:-502, 15 March 1957.

17. B. Wald, *Rapid Scanning Techniques for Countermeasures Receivers*, NRL Report 4410, September 1954.

18. P. I. Richards, "Universal Optimum-Response Curves for Arbitrarily Coupled Resonators," *Proc. IRE*, Vol. 34, p. 624 (1946).

19. L. Addleman, *Descriptive Specification EDL Search Receiver Set Type EDL-R1*, Sylvania EDL Report EDL-M13, 20 January, 1955.

20. D. J. Grace et. al., *The S-152—An Electronically Scanned Microwave Superheterodyne Receiver*, Report No. 152-3, Contract DA 36(039)-sc-63189, Stanford Electronics Laboratories, 20 December 1957.

21. *Receiver, Radio AN/ALR-7(VA-1)*, Raytheon Manufacturing Company, Santa Barbara Laboratory, August 1958, Contract AF 33(600)28015.

22. *Engineering Investigtion of Image Suppression for Backward Wave Oscillator Tuned Superheterodyne Receivers*, Final Engineering Report, Project 4589, Task 45035, Contract AF 30(602)-1663, Microwave Engineering Laboratories, Inc.

23. W. Batten et al., *Response of a Panoramic Receiver to CW and Pulsed Signals*, Tech. Report No. 3, EDG, The University of Michigan, June 1955.

24. W. R. Kincheloe, Jr., *The S-601 Micro-Sweep Receiver*, Report No. 601-1, Stanford Electronics Laboratories, 25 June 1958, Contract DA 36(039) sc-73151.

25. R. Kronert, "Pulse Compression," *NTZ-Nachrtech. Z.*, pp. 14-152, April 1957.

26. W. R. Kincheloe, Jr., *Frequency Measurement with Sweeping Receivers*, Stanford Electronics Laboratories Report No. 557-2, 1960.

27. R. Kronert, "Pulse Compression, Part II," *NTZ-Nachrtech. Z.*, pp. 305-308, July 1957.

28. R. E. Williams, "The Ratio-Sweep Intercept Receiver," *Transactions, Department of Defense Symposium on Electronic Countermeasures*, pp. 28-34, The University of Michigan, November 1957. (SECRET)

29. *MSJ-Multiple Signal Jammer Final Engineering Report*, p. 36, G. E. Advanced Electronics Center at Cornell University, March 1956 (Contract AF 33(616)-118). (SECRET)

30. Crane and Gilkey, *A Wide-Open Instantaneous Frequency Discriminator*, Report No. 518-1, Stanford Electronics Laboratory, to be published in April 1960.

31. J. Hadley, "Assessment of Systems for Use in an Airborne Frequency Determining Receiver," Ref. No. SL-294, Applied Electronics Laboratories of G.E.C., Ltd. (Great Britain), 3 August 1959.

32. J. K. Pulfer, *Microwave Receivers Using a Superregenerative Backward-wave Oscillator*, National Research Council of Canada, Report ERA-338, September 1957.

33. J. K. Pulfer, "Application of a Backward-Wave Amplifier to Microwave Autodyne Reception," *IRE Trans. on Microwave Theory and Techniques*, page 356, July 1959.

34. J. J. Younger et al., "Parametric Amplifiers as Super-Regenerative Detectors," letter to the editor of *Proc. IRE*, July 1959.

35. R. C. Cumming and J. E. Sterrett, *Antenna Noise Temperature Measurements at X-Band*, Report No. 703-1, Stanford Electronics Laboratories, 1960.

36. W. R. Kincheloe, Jr., *Field Intensity Measurements of AN/CPS-6B Radar Antenna Patterns*, Stanford Electronics Laboratory MTR No. 24, 30 December 1954.

37. R. E. Lee, *A Study of Search-Radar Vulnerability to Side-Lobe Jamming*, Stanford Electronics Laboratory TR No. 802-1, 22 April 1957.

10

Direction Finding

L. A. deROSA

10.1 Historical Data and General Principles

10.1.1 Introduction

The location of a radio-frequency source, whether it be a radar, a communication transmitter, or a navigational aid, is of importance to the determination of the enemy's strategy.

The activity in certain geographical areas may indicate an elaborate preparation under way; the activit   a single source at sea may indicate the location of a submarine; and he solution of a number of signal sources, synchronized to a common signal, may indicate a new air-defense function of extreme importance to our strategic bombing plans.

The determination of the geographic location of a signal source is in most cases accomplished by determining the intersection of the lines of direction from two or more spatially separating receiving points. Special cases not included in this general category will be discussed later.

Figure 10-1 indicates the elementary direction-finding procedure. The operator at a direction-finding installation A determines the position of a transmitter to be on a line AA', θ_A degrees with respect to North; this angle is known as the bearing of the signal source with respect to the receiver.

If, now, a second direction-finding station located at position B determines a line of direction BB' at a bearing θ_B to the signal source, the intersection D of the lines AA' and BB' locates the position of the signal source.
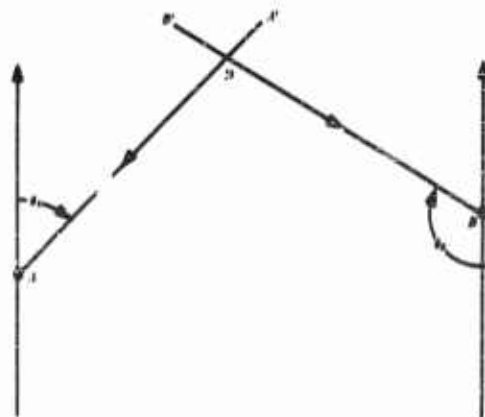
FIGURE 10-1. Location of a signal source *D* by direction finders *A* and *B*.

The accuracy with which a bearing can be obtained is influenced by three main factors:

(1) The accuracy of the direction-finding instrumentation including local site errors.

(2) The inaccuracies introduced by extraneous signals and noise sources whose direction of arrival is different from the desired signal and which are not separable by the direction-finding apparatus.

(3) The introduction of errors due to propagation anomalies, and the superposition of either discrete or diffuse images of the signal source thus produced at the observation point, which have other apparent directions of arrival than the original signal.

Methods reducing the errors from these various sources will be described later.

10.1.2 Historical Review

Prior to 1893, Hertz utilized, in his early experiments, cylindrical parabolic mirrors to focus and concentrate energy from a transmitter. His early tests used a frequency of about 200 Mc. Directivity was also used by Marconi before the turn of the century to increase the range of his transmissions. In his tests, copper parabolic mirrors were used to increase the range to 2 miles.

Since radio communication, because of the greater ranges possible, progressed rapidly to the hf range by 1900, the mirror techniques were no longer practical and J. Zenneck at about that time introduced simple directors and reflector wires, not simultaneously, but singly, at a little more than $\lambda/4$ to increase the range and to separate interfering signals.

The simple limaçon patterns of Zenneck progressed to large spacings of

λ/2 apart in the configurations of S. G. Brown, A. Blondel, and J. Stone. Stone, in 1902, proposed to physically rotate the array of Figure 10-2 to determine the direction of arrival of the wave. This will be recognized as one pair of an Adcock array.

Bellini and Tosi then followed by tilting the top of the vertical members of Figure 10-2 so as to facilitate the support by a common mast, and by introducing a phase shift in each element, as in Figure 10-3, were able to
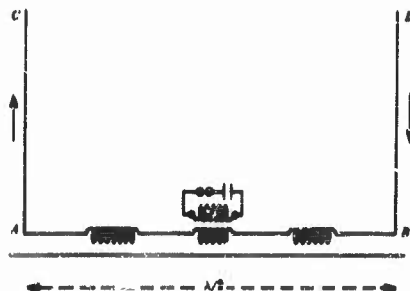
Fig. 10-2. The simple directional array
(Brown, Blondel, and Stone).
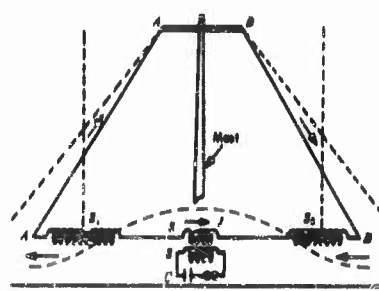
Fig. 10-3. The inclined-pair, single-
mast array of Bellini and Tosi.

reduce the spacing and yet obtain a null for a direction of arrival in the plane of the vertical elements. Then, in order to avoid the necessity for rotating the array, they provided a second array at right angles to the first and, by means of quadrature coils placed outside of a single rotating coil, also suggested by Artom, were able to sequentially sample the signal being picked up by each antenna pair. The device for sampling the antennas sequentially is called a goniometer. Such a device is shown schematically in Figure 10-4, the moving coil being connected to the receiver.
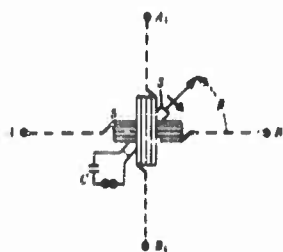
Fig. 10-4. The Bellini-Tosi gon-
iometer for sequentially sampling
a pair of the Bellini-Tosi array.

The Bellini and Tosi system proved quite successful in both transmitting and receiving systems in an installation at Boulogne where ranges of 1500 km were obtained with 500 watts of power at 10 Mc.

The next obvious improvement was the use by Zenneck of an omnidirectional element placed in the center of the array and so arranged as to add phasewise to the output of the moving coil in a Bellini-Tosi goniometer. This arrangement produced an antenna pattern which was a rotating cardioid rather than a figure-eight, thus eliminating the 180° ambiguity which existed with the latter configuration.

Just prior to this, in 1905 and 1906, Marconi was working with directive

antennas having a short vertical element connected to a long horizontal portion such as that shown in Figure 10-5, the directivity of which was shown to be as indicated in Figure 10-6.
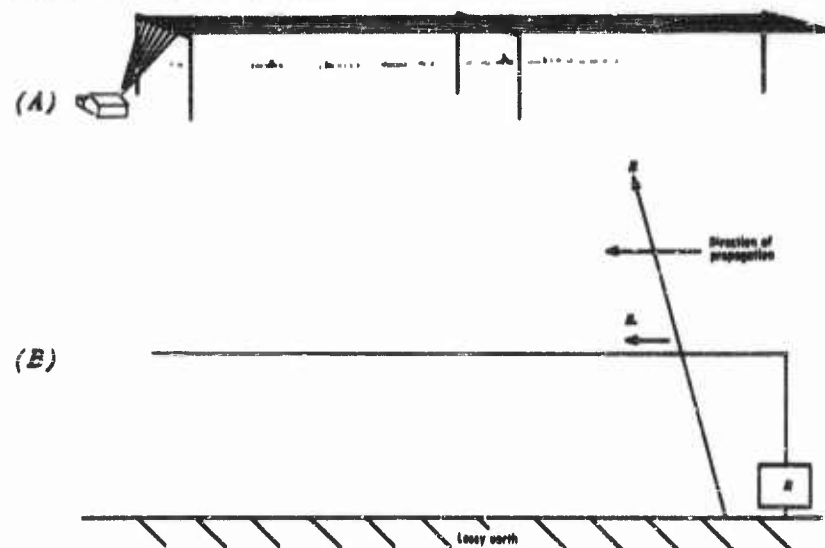
FIG. 10-5. The Marconi antenna. (a) A typical installation. (b) principle of operation, depending on lossy earth to tilt the electric field E.

Why such an array displayed directivity was a mystery for some time until H. von Hoerschelmann showed that the ground conductivity produced a tilt in the electric field, and that the horizontal component of this combined with the pick-up of the vertical section to increase the gain in the direction away from the feed point.

These so-called ground antennas were arranged by Marconi in radial configuration such as shown in Figure 10-7, each of which could be switched sequentially to the input of a receiver to determine the signal-arrival direction. Marconi obtained bearings from shipborne transmitters up to 90 km.
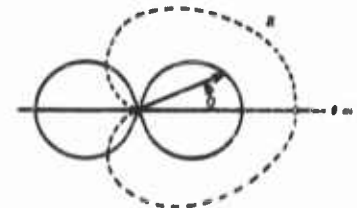
FIG. 10-6. Directivity pattern of the Marconi antenna. Solid curve is directional characteristic of the horizontal member. Dotted curve is the combined horizontal and vertical members.

$$E = K(1 + \cos\theta) = \begin{cases} 2K, & \theta = 0 \\ K, & \theta = \pi/2 \\ 0, & \theta = \pi \end{cases}$$
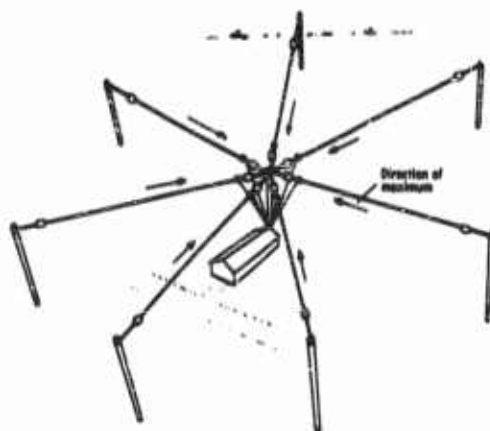
Fig. 10-7. Radial configuration of Marconi antenna for 360° coverage.

One of the first uses of a loop antenna for directivity tests was that of Hertz, who showed that, when a loop of wire feeding a spark gap was held in certain positions with respect to a radiating source, a spark would occur across the gap; whereas, for different orientations of the loop, this did not occur.

In 1905 and 1906, H. J. Round in New York described the use of frame antennas for directivity effects.

The First World War (1914-1918) produced many advances in direction-finding techniques, the vacuum-tube amplifier producing more sensitive receivers and test equipment, which accelerated the development of new types such as the Adcock, various loop direction finders, and numerous indicators, both mechanical and cathode-ray types.

The Second World War produced another period of rapid development in the direction-finding art. Crossed loops, eight-element Adcocks, Watson-Watt instantaneous, H collectors, sector-scanning, phase comparison or interferometer, lens, and many other improved types were developed and placed in military and civilian operation.

10.1.3 Basic Information Associated with Field Vectors

The numerous configurations of direction-finding antennas and systems derived, as many have been, by experimental methods, are apt to be confusing to one who follows the developments in chronological order. Accordingly, it might be well to consider the basic direction-finding problem from an informational-theory viewpoint, catalog old systems, and provide places in the catalog for the development of new systems.

In order to do this, the fields associated with an electromagnetic disturb-
ance should be considered and the information conveyed by the field vectors
relative to the signal source analyzed. All direction-finding arrays then can
be shown to be means for encoding this information in such a way as to
either simplify instrumentation or to mitigate the effects of error-producing
factors such as poor siting, propagation anomalies, and the deleterious effects
of extraneous signal and noise backgrounds.

The field at a direction-finding location can be regarded as having been
initiated by an rf source exciting an antenna. For a general discussion, the
elementary dipole can be considered as the basic building block for both the
transmitting and receiving antennas. The elementary dipole consists of a
perfectly conducting wire with a radius vanishingly small and of short length
so that the current in it can be of constant magnitude and phase throughout
its length. This theory can be applied also to a loop less than 1/10 wave-
length in circumference. To approximate larger antennas, they can be con-
sidered as a large number of infinitesimal or elementary dipoles with various
space distributions, polarization, current magnitude, and phase relations to
each other. The patterns of large antennas can then be obtained by summing
the field vectors of the many elementary dipoles.

If only a single electric dipole is considered, the geometry may be shown
as in Figure 10-8, where the electric and magnetic field components are
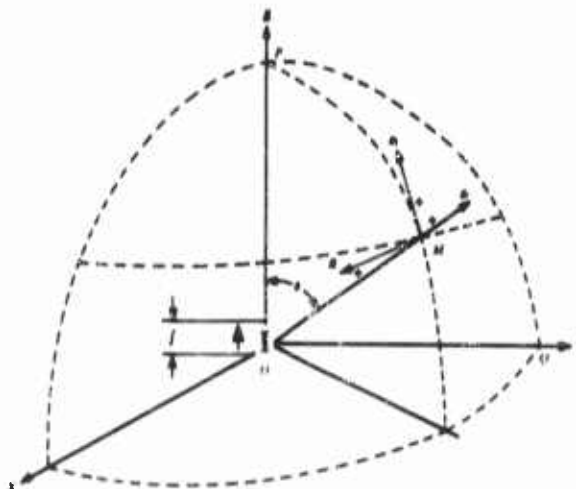
Fig. 10-8. Spherical coordinates for the dipole.

shown in spherical coordinates with positive values shown by the arrows. If
a magnetic dipole is to be used, the loop should be arranged perpendicular to

the s axis at the origin, in which case, the vector H becomes E, the electric field, E_t becomes the magnetic tangential field, and E_r becomes the radial magnetic field.

With the following representations:

$r =$ distance OM

$\theta =$ angle POM measured from P toward M

$I =$ current in dipole

$\lambda =$ wavelength

$f =$ frequency

$\omega = 2\pi f$

$\alpha = \dfrac{2\pi}{\lambda}$

$c =$ velocity of light

$v = \omega t - \alpha r$

$l =$ length of dipole

the equations for the electric and magnetic components are obtained in mks units.

$$E_r = - \frac{30 l \lambda I}{\pi} \frac{\cos \theta}{r^3} (\cos v - \alpha r \sin v) \tag{10-1}$$

$$E_t = + \frac{30 l \lambda I}{2\pi} \frac{\sin \theta}{r^3} (\cos v - \alpha r \sin v - \alpha^2 r^2 \cos v) \tag{10-2}$$

$$H = + \frac{l i}{4\pi} \frac{\sin \theta}{r^2} (\sin v - \alpha r \cos v) \tag{10-3}$$

For distances beyond about 10 wavelengths the radial electric field becomes negligible and only the tangential electric (E_t) and magnetic fields (H) need be considered. Then,

$$E_r = 0$$

$$E_t = - \frac{60 \pi l I}{\lambda r} \sin \theta \cos (\omega t - \alpha r)$$

$$H = + \frac{E_t}{120\pi}$$

Assume, at least for the present, that the polarization of the source, its geographical position, excitation level, and directivity (phasing the excitation of the composite elementary dipoles) remain fixed. Assume also proper instrumentation to indicate a minimum perceptible field variation ($\triangle E_t / E$ or

$\triangle H/H$) of 1 percent. Under these conditions for the usual distances involved, and under free-space conditions at a distance of 100 miles, the dipole must be moved on the circumference of a circle 1 mile in diameter. At greater distances, assuming that Eq. (10-2) and (10-3) hold, the movement of the dipole must be correspondingly greater.

During this exploratory movement, the power of the transmitter must be highly stable, the exploratory equipment must maintain calibration, and no perturbations of the field may exist. Thus it is difficult to determine the direction of incoming radio waves by noting the changes in intensity produced by the movements of the receiver. However, in the case of fast-moving aircraft, "sniffing" operations may be feasible. They are discussed later in connection with the ASQ-23 equipment.

The attribute of the electric and magnetic field vectors other than their magnitude is their instantaneous direction as a function of the space coordinates X, Y, and Z.

In free space, or over a homogeneous ground plane, all points equidistant from an isotropic signal source will have the same instantaneous phase with respect to each other.

Although antenna arrays used in transmitters produce a signal whose phase varies around the array, at a distance and with the usual direction-finding array, the equiphase front is usually normal to the direction of propagation. At hf, even an antenna aperture of 100 wavelengths will produce an equiphase front at equidistant points for more than a half mile sector at 100 miles distance. Thus it is evident that, even for the largest aperture transmitting antenna, phase measurements to determine the phase front will be most effective.

Reflections from large objects close to the transmitter or the receiver, however, will distort the phase fronts, and special consideration must be given to those cases to obtain accurate bearings.

The field at a point remote from a transmitter has been found from Eq. (10-2) and (10-3) above to consist of electric field and magnetic field vectors represented by E_i and H.

These vectors not only determine the amplitude but the instantaneous phase of the field with respect to either the field at any other point, or to a stored signal having the exact same frequency at all times as that represented by the field vectors.

Under these conditions, it is possible to consider the direction-finding problem as one of exploring either simultaneously, or sequentially, the field vectors at various closely grouped points to determine the equiphase front and thus determine the direction of arrival of the signal.

The information contained in the signal is limited to, and does not exceed,

that supplied by the field equations. As was shown above, the coefficients of the field equations vary so slowly as a function of distance that only the phase of the field vectors is useful.

10.1.4 Array Classification

The determination of the direction of arrival of a radio wave may be accomplished by a number of simultaneous or sequential measurements of the phases of the field vectors at different points. The selection of the sampling points, and the manner in which outputs of the exploring elements of the direction finder array are combined, may be considered as an encoding process designed to either simplify the instrumentation and/or to reduce the error due to local site disturbances.

For simplicity, the various exploring techniques can be classified by the following grouping:

Absolute or instantaneous methods of measurement of

    (a) Amplitude $(A) = f(\theta)$
    (b) Phase $(\phi) = f(\theta)$
    (c) Delay $(\delta) = f(\theta)$

and sequential methods of measurements of

    (a) Amplitude $\partial A/\partial\theta = f(\theta)$
    (b) Phase $\partial\phi/\partial\theta = f(\theta)$
    (c) Delay $\partial\delta/\partial\theta = f(\theta)$

where θ is the azimuth angle of the array with respect to a reference, generally North.

As an example of the encodings represented by the methods above, a few configurations are considered.

To determine the direction of arrival instantaneously by amplitude methods, a pair of antenna elements may be connected together to form a figure-eight pattern, as shown in Figure 10-?(a). A similar array may be arranged at right angles, as shown in Figure 10-9 (b).

Each of the two arrays may be considered as converting the phase field into an amplitude function of azimuth, the amplitude being a sinusoidal function of the angle between the plane of the antennas and the equiphase front; that is,

$$A_1 = K \sin \theta$$

and

$$A_2 = K \sin (\theta + 90°) = K \cos \theta$$

(a)

$A_1 = K \sin \theta$ $A_2 = K \cos \theta$

(b)
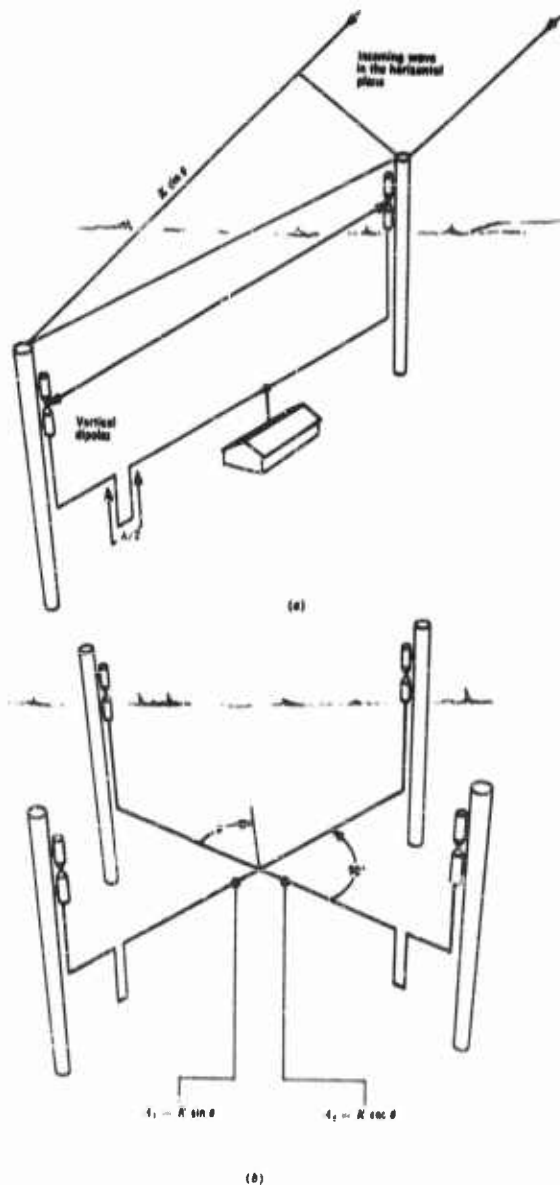
FIG. 10-9. (a) Antenna array having a figure-eight pattern. $A_1 = 2 \sin [(K/2) \sin \theta]$ for $K \ll 2\pi$; $A_1 = K \sin \theta$. (b) Two figure-eight arrays at right angles.

where K is the spacing between antenna elements in electrical degrees at the frequency of the signal.

The direction of arrival is obtained by a comparison of the outputs of each antenna pair. Thus,

$$A_1 = K \sin \theta$$

$$A_2 = K \cos \theta$$

or the ratio

$$\frac{A_1}{A_2} = \frac{\sin \theta}{\cos \theta} = \tan \theta$$

or

$$\theta = \tan^{-1} \frac{A_1}{A_2}$$

As an example of a phase method of measurement for instantaneous indication, assume two probes which do not disturb the field; from the outputs of these, a Lissajous figure can be formed by connecting each probe to a set of deflecting plates of a cathode-ray oscilloscope. If the frequency of the signal and the spacing between the probes is known, then

$$\phi_1 - \phi_2 = \frac{2\pi d f}{v} \sin \theta$$

or

$$\theta = \sin^{-1} \frac{(\phi_1 - \phi_2) v}{2\pi d f}$$

where $v =$ velocity of light
           $d =$ distance between probes
           $f =$ frequency of signal
    $\phi_1 - \phi_2 =$ difference in rf phase between point 1 and point 2

The ambiguity due to symmetry about the horizontal axis can be resolved by the use of a third probe.

At this point it may be well to reiterate the note that all methods utilize the phase of the field vectors. Some encode by transforming the phase rela-

tions to an amplitude function, while others compare the phase of the field vectors at two or more points, either simultaneously or sequentially.

An example of the use of a delay method of determining direction is shown in Figure 10-10. Here the signal is rectified and the latency in the time of arrival at two points is measured.

$$t = (d \sin \theta)/v \qquad \text{or} \qquad \theta = \sin^{-1} vt/d$$

where d = distance between probing points
v = velocity of light
t = time difference between the arrival of a reference point and the modulated wave at point 1 and point 2

Ambiguities may be avoided by the use of a third probe. This method is often referred to as the "inverse loran" system. If the radio wave is un-
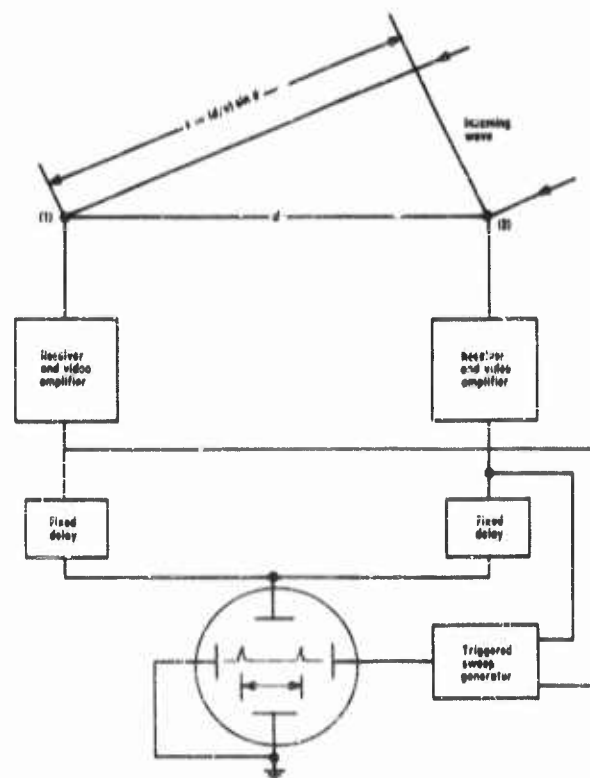
Fig. 10-10.   Time-delay method or inverse loran system of direction finding.

modulated, that is, has no appreciable sidebands, this degenerates to a phase-measurement case. The latency in time of arrival may be measured by a number of methods and will be discussed later.

10.1.5 Scanning Types

Antenna arrays for direction finding may be simplified by probing the vector field in a sequential manner using the same antenna, but moving it from point to point and noting the manner in which the phase of the field changes.

In some cases more than one fixed antenna array may be used and each antenna connected in sequence to a receiver. The sequential-probing, or as they are often called, scanning methods may often offer advantages when the signal is stable and of long duration relative to the scanning cycle.

An example of a sequential-probing method encoded so as to produce an amplitude change as the vector is explored is the rotating cardioid type shown in Figure 10-11.
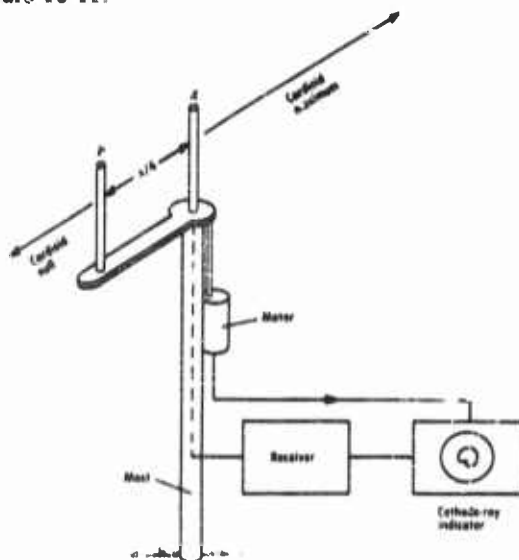
Fig. 10-11. System for probing the field, using a rotating cardioid.

In this type, an antenna A is fixed on a mast and a parasitic element P is rotated around the antenna element A at a distance slightly less than $\lambda/4$ at mean frequency. The antenna A probes the field at a fixed point while the parasitic element probes the vector field on the circumference of a circle $\lambda/4$ radius from A. The field probed by element P is coupled to A by the mutual

coupling so that the output of antenna A is a function of the vector fields existing at the center and on the circumference of the circle, and thus a change in amplitude is produced by the antenna rotation. The amplitude change is substantially sinusoidal with a period of one revolution, so that the phase of the envelope of the radio wave thus detected with respect to a fixed reference indicates the direction of arrival of the radio wave.

The analysis of such a direction-finding system may also be considered simply as a directive array which is rotated, and the direction of the transmitter is determined by noting the orientation of the array when the signal is maximal.

The minimum of the directive pattern may be used in a similar fashion, but the information contained in the probing of the complete circle is partially lost by such a restricted search, particularly when the signal level is low.

10.1.6 Sequential Phase Method (Doppler Direction Finding)

If an antenna is moved at a uniform speed across a vector field, if the movement is not tangential to the equiphase contours, and if the frequency of the signal is known, then the movement necessary to produce 2π radians of phase change may be experimentally determined. This 2π radians or $360°$ is equivalent to producing a 1-cycle frequency change during the time required for the movement. The azimuth of the transmitter may then be determined as shown in Figure 10-12 and is simply:

$$\theta = \cos^{-1}(\lambda/d)$$

where λ = wavelength of the signal

d = distance moved in a straight line to change the phase of 360 electrical degrees with reference to an oscillator having the identical frequency as the received signal

If the antenna is moved on the circumference of a circle, then the phase change, which is equivalent to introducing a doppler frequency shift, can be measured; and if the velocity of movement is constant, the direction of the transmitter may be determined by noting the position at which the doppler shift is maximal; the line of direction is tangent to the circle at this point and in the direction of movement of the antenna.

Many variations of sequential phase measurement are possible. The antenna need not move; the phase at different fixed antennas positioned in a circle can be sequentially compared and the result will be essentially equivalent.
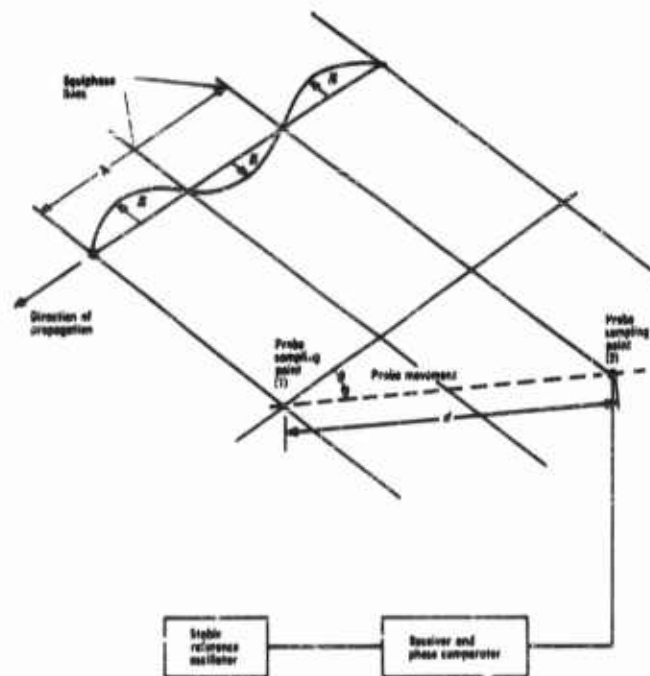
Fig. 10-12. Sequential phase method of direction finding.

10.1.7 Sequential Time-Delay Methods

The latency in time of arrival of a characteristic feature of modulation in a wave can be determined at two locations and by moving one position around the other. The relative positions for simultaneous arrival can be noted, and from this information the direction of arrival of the signal is found to be normal to the line connected to these positions. The 180° ambiguity can be resolved by a movement of one antenna along the line of direction.

10.1.8 General Considerations Determining the Choice of Techniques

In many cases, as will be evident in the following discussion, no one type or method will be satisfactory, in which case the techniques may be combined as a system. Take as an example an occasion when both short-burst transmissions (fraction of a second) and longer cw-type signals must be received. Although the sequential or scanning direction-finding techniques may be more accurate and more easily instrumented, a short transmission will

not permit sequential probing; therefore the instantaneous type technique must be used. However, a sequential phase or doppler type may be used at one position and a separate antenna provided many miles away using inverse loran techniques.

The factors that determine the choice of a system include, generally, permissible size of the array, operating frequencies, terrain characteristics, frequency band to be covered, polarization of the incoming signal, signal duration, bandwidth of the signal, accuracy of bearing required, and data-processing equipment available.

These factors will be discussed later, but in general we may regard the electromagnetic field and the associated vectors as an ideal field perturbed by a variety of irregularities. These irregularities are due to propagation through a nonhomogeneous medium that is bounded by a ground plane neither smooth nor electrically uniform. In the case of waves reflected from the ionosphere, the upper boundary is not only nonuniform electrically but changing with sunspot activity, time of day, season, and latitude. In addition, the boundaries and at some points even the medium between the bounds may be dispersive; that is, its propagation characteristics may vary with the frequency of the signal.

Any factor which tends to distort the phase front of a signal will produce an error in the direction-finding procedure; the accuracy with which a line of direction is obtained is dependent on a knowledge of the distortions which may exist and in the proper selection of the direction-finding techniques to minimize the errors which these distortions produce.

The field vectors represent a space- and time-variant signal with certain imposed constraints and with error-producing "noise" disturbances.

If simultaneous probing is used in exploring the field, the data obtained at each exploring point are not independent for the case where power is extracted from the field. The data thus obtained are subject to an error due to the mutual coupling to the other elements of the array. The reduction of these errors will be considered later.

If a single probe is moved sequentially from one exploring point to another, the mutual coupling problem is replaced by the requirement that the transmitter and receiver be stationary, and the frequency stability of the transmitter be of a high order. Attempts to probe at rapid speeds to avoid errors from frequency changes introduce sideband frequencies, necessitating wider bandwidths with the associated deteriorations in signal to noise ratios.

The reference to the role of channel capacity required to offset an uncertainty in the frequency of the transmitter, and the analogous minimum sampling intervals due to mutual effects lead to many interesting fundamental relations.

The selection of an exploring path, or the interconnections of a series of antenna elements in an array, may be regarded as an encoding procedure to reduce the errors due to disturbances. As in information theory, it is generally found that the use of *a priori* data relative to the transmitter or the propagation medium results in a lower error rate, smaller integration times, and simpler instrumentation.

10.2 Propagation of Electromagnetic Waves

The accuracy of a DF (direction-finding) equipment is influenced by the signal characteristics, design of the DF apparatus and antenna, its local siting conditions, the data-processing methods used, and the characteristics of the propagation medium.

The apparent direction of arrival of a signal is determined finally by the irregularities of the intervening medium. The errors introduced by various factors associated with the propagation of the electromagnetic wave impose practical limits on the range and accuracy of the DF operation.

A brief discussion of the uncertainties of propagation is helpful in indicating the relationship between residual errors from that cause, and the methods, by the use of more complex equipment and procedures, for reducing such errors.

The errors caused by propagation effects will be discussed here, while the reduction of such errors will be treated later (Section 10.9). The rf frequencies of interest to the reconnaissance operator range from the ELF (extremely low-frequency) waves (10-500 cycles) to the microwave region beyond 125,000 Mc.

Through this wide frequency scale, over 10 decades, the factors which introduce disturbances in the path of propagation are many and are related to the frequency of the signal.

Within the bounds of the earth and the ionosphere, inhomogeneities are found due to changes in the electromagnetic properties of the earth, geometric irregularities of the surface, nonuniform refractive and absorptive properties from various causes in the troposphere (0-12 km height), in the stratophere (12-80 km height), and in the ionosphere (80-400 km height).

If the propagation path extends beyond the ionosphere, as in satellite communication and radio astronomy, except for the unusual case of diffraction in the immediate neighborhood of celestial bodies, substantial deviations from linear propagation will probably occur only as a result of transmissions in or through the ionosphere.

To the normal ionosphere inhomogeneities must be added those produced by unusual solar activity and nuclear blasts.

10.2.1 Electromagnetic Properties of the Upper Atmosphere

The transmission of radio waves beyond the horizon is due primarily to the phenomenon of diffraction caused by geometric irregularities of the earth's surface; transmission by surface waves which follow ducts created by striations of media having different dielectric properties; refraction and scattering from tropospheric areas, and the reflection and refraction from ionospheric layers.

Of these, the ionization in the upper atmosphere (from 60 to 400 km height), of the gases present there, is an important consideration in the propagation of radio waves over long distances beyond line-of-sight limitations.

The ionization in this region is not uniform but is arranged in "layers" of greatest ionization. Figure 10-13 shows the location of these ionized layers.

At times another region E_2 is formed above E. The layer shown as D is essentially an absorbing layer and disappears at night. The F_1 and F_2 layers merge at night to form a single layer F. The four main regions of ionization and D, E, F_1 and F_2. The maxima for the F_1 and F_2 layers are located at 200 and 275 km, respectively. The daylight layer D lies between 60 and 100 km. While it is essentially an absorbing belt, nevertheless at the lowest frequencies it also aids in the propagation of radio waves.

The ionizations produced by ultraviolet radiation, and corpuscles from the sun hence vary daily, seasonally, and with magnetic storm and auroral activity.

Fig. 10-13. Ionized regions of the upper atmosphere.

Ionization is also produced by the passage of meteors, satellites, and missiles.

10.2.2 Electromagnetic Waves in an Ionized Medium

If the ionized medium contains free electrons, then the movement of these electrons is affected by the passage of an electromagnetic wave. The electrons will vibrate and produce a current which is added to the displacement current due to the electromagnetic field itself.

The net result, neglecting collisions of electrons with neighboring neutral

atoms and molecules and also the effect of the existing magnetic field, is to reduce the dielectric constant of the medium by the quantity

$$4\pi N_e e^2/m\omega^2$$

where N_e is the number of electrons per cubic centimeter, each having a charge e and a mass m, and ω is the angular frequency of the wave $2\pi f$.

In empty space, K, the dielectric constant of the medium, is equal to unity, so that the presence of the electrons reduces it from unity to

$$1 - (4\pi N_e e^2/m\omega^2)$$

The phase velocity of the electromagnetic wave in the medium is therefore changed from a value c to u, where

$$u = \frac{c}{\sqrt{1 - (4\pi N_e e^2/m\omega^2)}}$$

This reduction of the dielectric constant to a value of less than unity, and the associated increase in phase velocity, produce a bending downward of electromagnetic waves transmitted upward, as shown in Figure 10-14.

Fig. 10-14. Bending of an electromagnetic wave by an ionized layer. The upper portion of the wave (A of AB) moves faster since it is in a region of higher electron density.

If the electron density at any given height changes with geographical location, then reflections from ionized layers will also be laterally diverted; that is, the signal will not arrive at the receiving point from a direction of the transmitter.

Besides this lateral deviation, irregular densities of ionization will produce, at any given receiving point, a signal intensity due to the contributions (in or out of phase) of a plurality of paths.

Since the electron densities in the various ionized layers are variable with time, being influenced by drifts, winds, and turbulence, the reception may be highly variable; and under certain conditions the apparent direction of arrival will fluctuate violently.

Fluctuations will occur in particular during periods of magnetic storms, auroral activity, and when the sunset and sunrise lines are in somewhat the same direction as the direction of propagation.

10.2.3 Absorption in the Ionized Layers

If we consider the effect of collisions of the excited electrons with neutral particles and heavy ions and other electrons, it is found that energy is extracted from the electromagnetic wave.

The ionosphere may be considered as having a conductivity of

$$\sigma = \frac{N_e e^2 \nu}{m(\omega^2 + \nu^2)}$$

where ν is the frequency of collisions per second, approximately 10^6 for the E-layer and 5×10^3 for the F-layer. The dielectric constant therefore is

$$K' = K - j\,\frac{\sigma}{\omega \epsilon_0}$$

and the complex refractive index M is

$$M^2 = K' = \left(\mu - \frac{jck}{\omega} \right)^2$$

where μ is the refractive index for the medium and k is the absorption per unit length of path (Reference 1).

The absorption is of interest in DF techniques in determining the relative transparency of an ionized layer, particularly as it affects waves incident at planes not normal to the layer.

10.2.4 Propagation in a Magnetic Field

The ionosphere is within the earth's magnetic field, and the reflection of waves under these conditions produces splitting of the wave and polarization changes, both of which may affect the observed direction of arrival of the signal.

When a magnetic field is present, a single incident wave will be split into two waves, the ordinary and the extraordinary. This process is similar to double-refraction phenomena in optics. The wave which is less affected by the magnetic field is called the ordinary one.

There are two conditions depending on whether the gyro frequency

$$f_g = eB/m$$

of the electron in the earth's magnetic field is less or greater than ω, the angular frequency of the incident radio wave. In the above formula B is the

earth's magnetic induction field in webers per square meter, e is the electron charge, and m is the electron mass.

When the gyro frequency eB/m is less than ω, there are three electron densities (which normally occur at three different heights) for which reflection can occur. These are obtained by letting

$$\frac{4\pi N_e e^2}{m\omega^2} = 1 - \frac{eB}{m\omega} \tag{10-4}$$

$$= 1 \tag{10-5}$$

$$= 1 + \frac{eB}{m\omega} \tag{10-6}$$

When the propagation of the wave is in the direction of the magnetic field, only Eq. (10-4) and (10-6) hold.

When the gyro frequency eB/m is greater than ω, there are two densities for reflection. These are obtained by letting

$$\frac{4\pi N_e e^2}{m\omega^2} = 1 \tag{10-7}$$

$$= 1 + \frac{eB}{m\omega} \tag{10-8}$$

When the propagation of the wave is in the direction of the magnetic field, then only Eq. (10-8) holds.

The computations above hold for the condition when absorption is neglected.

10.2.5 Polarization Changes

The magnetically split waves are elliptically polarized in opposite angles of rotation as they enter the ionosphere. The polarization ellipses are of the same eccentricity but at right angles to each other (page 667 of Reference 2). For the northern hemisphere, the component having the counterclockwise rotation will be reflected where

$$\frac{4\pi N_e e^2}{m\omega^2} = 1 \pm \frac{eB}{m\omega}$$

In general, however, reflection occurs at the lower level defined by the equation

$$\frac{4\pi N_e e^2}{m\omega^2} = 1 - \frac{eB}{m\omega}$$

and therefore the wave does not reach the height necessary to produce a plane polarization condition.

The sense of polarization is reversed on reflection, an upward-going wave with a counterclockwise rotation being converted to a downward-going clockwise polarization. The polarization diagram is rotated by the effects of absorption due to collisions of excited electrons.

10.2.6 Meteor Trails

Transient discontinuities in the ionosphere produced by meteors may produce deviations in the direction of arrival of a signal.

The increases in E-layer ionization at night during meteor showers have been fairly well correlated; in fact, observations of the critical frequency at night during the maximum period of the Leonid meteor shower of 1932 indicated electron densities greater than the noon value for a summer day (Reference 3).

Radio signals have been reported over long distances at frequencies where both the E- and F-layers are ordinarily penetrated (Reference 4). Signals at frequencies as high as 40 Mc have been observed at long distances. Signals received from such ionized patches will obviously have an apparent direction of arrival determined by azimuth of the disturbance.

10.2.7 Nuclear Blasts

An atomic blast at high altitudes will ionize the atmosphere, thereby either greatly enhancing the effect of the ionosphere, or creating clouds of electrons.

At an altitude of 60 miles, a kiloton atomic bomb would produce a sphere of 5-km radius within which there would be an average electron density of 10^{12} electrons per cubic centimeter or about 10^6 times the average daytime density of the E-layer.

Because of the high densities, $N_e = 10^{12}$, reflection would occur up to the region of X-band (10,000 Mc).

In addition to this, beta particles would be carried by the magnetic field to the opposite hemisphere where, at heights of 50 to 100 km, electron densities of $2 \times 10^8/cm^3$ would be produced. The fission fragments, also guided by the earth's magnetic field, would arrive a few seconds later and produce, at higher altitudes (100-200 km), electron densities of $3 \times 10^9/cm^3$.

The low-level disturbances would mainly be absorbing, preventing long communication except at the very lowest frequencies, where, however, DF operations would be difficult because of the variation of the phase velocity and of the consequent severe phase-front distortions.

The reflecting patches of ionosphere at high level would therefore seem to be a virtual source for many signals of higher frequency, that is, above a few hundred megacycles.

10.2.8 Lower-Atmosphere Phenomena

Useful electromagnetic fields can be propagated several hundred miles at frequencies of about 40 to 4000 Mc. The received signal is relatively independent of frequency, but the signal strength may vary considerably, depending on the intervening weather and atmospheric conditions. When observing such signals, two types of fading are encountered. The first type has an amplitude following a Rayleigh distribution over short periods during which the troposphere can be considered constant. This fading, of a rapid type, is due to many paths of slightly different length and can be reduced by space diversity.

The second type of fading is much slower and is caused by variations in the gradient of the refractive index of the atmosphere; this type is not aided by diversity.

In lower-atmosphere propagation, any artificially introduced reflector such as an aircraft, particularly if flying close to the transmitter or receiver, will increase the field intensity by large amounts depending on the geometry. A bearing taken on the transmitter will, here, be that of the aircraft.

Similarly, if the transmitter has a sharp beam which is being rotated, the bearings will scintillate due to diffraction and reradiation from various high spots on the horizon for distances somewhat beyond line of sight, but will show less error at large distances. The resolution of the bearing will in general be poor and bearings will move randomly through large deviations, a phenomenon known as "galloping." Ranges to be expected from troposphere scatter are available (page 759 of Reference 2).

In addition to the relatively steady tropospheric scattering of signals, unusual gradients of defractive index have allowed the reception of radar returns at large distances, a range of 1700 miles having been reported during World War II. When the index of refraction decreases rapidly with height, high-frequency signals can be propagated as in a waveguide.

Favorable conditions for such duct propagation may be found when the air temperature is higher than that of the water or when two air masses of different temperatures are contiguous for a long distance. For an extensive treatment of phenomena of this sort, see Reference 5.

10.2.9 Lacunae

Propagation conditions such as ducting, scattering, and ionospheric propagation have been studied extensively because of their importance in in-

creasing the range of communication and radar-detection systems. In addition
to phenomena from these causes, the existence of lacunae, or islands of at-
mosphere having discontinuities of refractive index due to moisture, may be
encountered during the stages of cloud formation, and vertical ducting due
to nonuniform heating of terrain may occur on summer days when wind
velocities are low. These lacunae produce, in the vhf and higher frequencies,
"galloping," or rapid fluctuations of the bearing, particularly when the trans-
mitter source is close to the horizon and a land-based direction finder is
being used.

10.2.10 Earth Effects

The irregularities of terrain produce disturbances in the propagation of
radio waves and therefore difficulties in the DF procedures. The presence of
reflecting objects such as mountains disturbs the phase front and produces
interference effects. Mountains and cliffs also cause diffraction phenomena,
which are especially noticeable at vhf and above, depending on the topog-
raphy.

Differences in the earth's conductivity, particularly in the transition be-
tween sea water and ground, also give rise to interference phenomena and
disturbances in the phase front. The latter effect is particularly noticeable
at the hf and lower frequencies and is known as coast refraction (References
6, 7, and 8). It produces errors in the direction of arrival of 4° to 5°. The
level of the tide has been found to affect the direction of arrival in one case
at least by 1.25° for direction almost parallel to a coastline.

When a reflected wave undergoes a change in polarization, or when the
original signal contains a wave other than a pure plane-polarized wave, errors
will occur in the apparent direction of arrival since the DF antenna patterns
will often have a different azimuthal pattern for each polarization. More
will be said about this source of error as it is encountered in the discussion
of DF antenna arrangements.

10.2.11 Local Site Effects

The effects of local reflections on the accuracy of a DF equipment are
difficult to analyze mathematically because of the irregular shapes, orienta-
tions, and dielectric discontinuities which one encounters in local sites. Typical
examples are the problems associated with DF apparatus installed in aircraft
and operating at frequencies of about 3 to 300 Mc.

The treatment is further complicated by conversions of polarization and
the effect of such modified polarization on the DF array. Further local site
reflections, being in proximity to the array, will affect different elements of
the DF antenna nonuniformly. In general, except for certain exceptional cases,

the reradiation phenomena cannot be treated as Fraunhofer diffraction, but must consider the effect of phase differences in the wavefront which excites the antenna; therefore, the Fresnel diffraction case must be treated.

To indicate the character of the difficulty, an extremely simple reflector may be considered. Assume a metal sphere of radius r on which is impingent a plane wave of length λ. If the center of coordinates is chosen at the center of the sphere, it is found that the scattered field has radial components which fall off as $1/r^3$ and has transverse components which fall off as $1/r$. Hence for large values of r, the electric and magnetic fields vary as e^{jhr}/r, where $h = 2\pi/\lambda$ is the phase-propagation constant. The Poynting vector at large distances varies as $1/r^2$. If the time average of the Poynting vector is integrated over a sphere of large radius and is concentric with the scattering sphere, this yields as the total power reradiated from the scatterer:

$$-\frac{dW_s}{dt} = \frac{1}{2} R_e \int (E_s \times H_s{}^*) r^2 \sin\theta\, d\theta\, d\phi$$

where E_s and H_s are the components of the scattered field.

The power is the incident wave per unit area. If E_0 is the magnitude of the incident electric field, this power is

$$\frac{dW_1}{dt} = \frac{1}{2} E_0{}^2 \sqrt{\frac{\epsilon_0}{\mu_0}}$$

The ratio

$$\frac{dW_s/dt}{dW_1/dt} = \sigma$$

is by definition called "scattering cross section."

In general, the scattering cross section is a complicated function of $hr = 2\pi r/\lambda$. However, if $hr << 1$

$$\sigma = (10\pi/3) k^4 r^6$$

that is, the scattering cross section varies as the inverse fourth power of the wavelength, and is smaller than the geometrical cross section πr^2.

Let such a small metallic sphere be moved about in space at a distance d_1 from a remote transmitter and at a distance d_2 from a DF array, then the effect of the scattering sphere will be, at the receiver,

$$P(t) = Pe^{j\omega t} \cdot \sigma \left(\frac{1}{d_1 d_2} \right)^2$$

where ω is the angular frequency of the rf signal and P is a constant dependent on the transmitter characteristics and on the scatterer pattern.

For a given scatterer, and assuming that $d_1 + d_2$ is a constant, as either d_1 or d_2 approaches zero, the amplitude of the scattered signal at the receiver increases. The maximum phase disturbance is produced when the scattered signal appears at the receiving antenna in phase quadrature with the existing field. A scatterer with the above description, then, will have the most effect when it is close to the receiver and so placed as to introduce a signal in phase quadrature with the direct signal.

If a number of antennas are used as part of an array, then, unless an unusual coincidence prevails, the maximally disturbing phase will not occur at all antenna elements at the same time and the error will in general tend to average out.

Indeed the design of an antenna with a null in the direction of the reflecting object can be considered as equivalent to a selection of antenna element geometry, of the adjustment of their feed lines, and of the proper weighting of the coefficients for each element so as to cancel all phase and amplitude disturbances due to the single reflecting element.

If the diameter of the reflecting sphere increases, that is, if kr approaches or exceeds unity, the effect of the reflecting object is more difficult to analyze.

The reflecting sphere under these conditions is excited in higher modes and therefore has a nonsymmetrical reradiating pattern which may at some frequencies produce no disturbance at the DF site, but at other frequencies may have a maximum effect. This will also vary with the polarization of the exciting wave and, of course, with the positions of the reflecting object, of the transmitter, and of the receiver.

From the mathematical analysis for the large sphere, a similar attack can be extended to the very prolate sphere which is the solution for a wire antenna of finite length.

The case of irregular sheets of metal, hangers, and ship structures can only be estimated although attempts have been made to formulate mathematical models (Reference 9).

A reflecting object such as a flat metal sheet many wavelengths across produces a reflection which as indicated above is dependent on the orientation of the sheet. As the size is increased, the "edge effect" (the radiation due to excitation of the sheet near the edges and to conductive currents from other portions which have not yet been reduced to negligible values by the reradiation) becomes of less and less importance. Finally, for a large sheet

and for a wave at normal incidence, the reflection is such as would occur if an antenna of the same area were uniformly excited across its aperture. As the sheet is turned, the major lobe moves so that it is directed along a line such that the incident and reflected angles are the same as in optics. There will be energy reradiated by a minor-lobe structure. This minor-lobe structure depends on the linear edge dimensions, while the main-lobe structure is more nearly dependent on the area of the sheet. The main lobe will not change orientation as a function of frequency, although the minor lobes will change.

10.2.12 Ground Conductivity

In direction-finding equipment, variations in the dielectric constant and conductivity of the terrain in the vicinity of the antenna can cause appreciable errors unless corrected by swamping such variations by high-conductivity screening, counterpoises, or chemical treatment of the ground. The terrain should preferably be uniform for an area at least 3λ in diameter. It can be made so by a counterpoise consisting of wire mesh less than 0.1λ square.

A simple example to show the effects of irregular ground planes can be presented by considering that, for every direction of arrival, each element of the antenna must see a similar image of the transmitting antenna in the ground plane. As the antennas are lowered toward the ground plane, the necessity for uniformity becomes more important, although the area over which uniformity must be maintained may be smaller. Uniformity is most important for signals which are descending since the ground plane close to the antenna elements contributes most to the formation of the image under these conditions.

10.3 Sequential Amplitude-Measurement Types

In the sequential amplitude-measurement type of DF technique, a directive antenna pattern is moved in azimuth and the changes in amplitude are observed, the direction of arrival being determined by the azimuth at which some amplitude criterion is met.

The most common type of antenna used is the loop shown in Figure 10-15(a). Such a circular loop carrying a uniform current distribution and with a diameter d, small with respect to the wavelength λ of the signal, can be shown (Figure 10-15b) to have a figure-eight pattern in the plane of the electric vector—that is, in a plane normal to any two parallel sides. As the size of the loop is increased, higher-order odd harmonics with respect to azimuth angle θ are introduced until, for a diameter of λ, the pattern changes to that of Figure 10-15(c) and for the 5λ-diameter case to that of Figure 10-15(d).
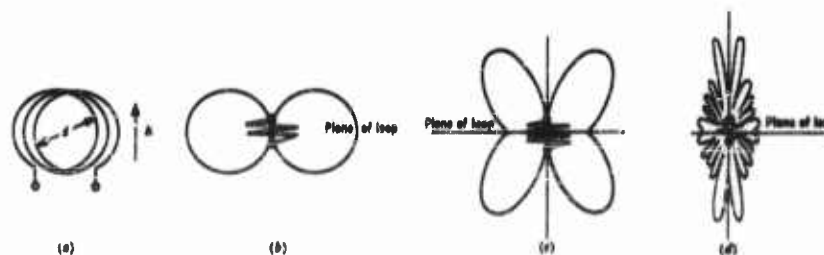
Fig. 10-15.   Far field patterns of loops 0.1, 1, and 5 λ in diameter with uniform in-phase current. (b) $d = \lambda/10$ (c) $d = \lambda$. (d) $d = 5 \lambda$.

If the loop were arranged so that its plane were normal to the ground, then for a signal arriving parallel to the ground—that is, with an angle of elevation equal to zero—the antenna pattern would appear as a figure-eight with maximum the plane of the loop. This would hold only for a vertically polarized signal, since there would be no response to a horizontally polarized signal, nor, of course, to any horizontal component of a complex polarization.

If, on the other hand, the signal being received were downcoming and contained a horizontally polarized component, then the loop would show a maximum response to the horizontally polarized component in space quadrature with the maximum for the vertically polarized component. The result would be to produce an error which would tend toward 90° of azimuth as the polarization became entirely horizontal. The error is also dependent on the relative time phase of the horizontal to the vertical component of the incoming signal. For standard test conditions for polarization errors, the IRE standards may be consulted (Reference 10). Polarization errors are further discussed in Section 10.9.

In practice, and in its most elementary case, a loop is rotated until the audio output (if applicable) is either a maximum or a minimum. When the signal is at a maximum, the loop is pointing in the direction of the signal (that is, its vertical sides are aligned in the direction of the signal). If a minimum is sought, then the direction of arrival is broadside to the loop. It is generally advisable to seek a null, particularly when the signal-to-noise ratio is good, since the antenna pattern is changing most rapidly at that point. Polar diagrams, however, give an improper indication of the true sharpness of antenna patterns, and it is sometimes better to consider a plot in rectangular coordinates.

The greatest rate of amplitude change with azimuth occurs at the null position, although in listening to a demodulated signal, the factors which affect the discrimination of the null are far more complicated and must include a consideration of the percentage of modulation, the detector character-

istic, noise intermodulation effects, and the logarithmic response of the ear. Further, many varieties of nonlinear amplitude circuit characteristics may be used to produce an apparent sharpening of the antenna pattern.

10.3.1 Shielded Loops

The loop can be operated as an electric dipole (the so-called monopole) by connecting a receiver to either or both ends and to ground, either directly or through a capacitor or an inductor. In this case, the pattern will, in general, be omnidirectional for vertical polarization except for possible irregular images in the ground plane. This mode of operation of the loop distorts the loop pattern, which may shift and fill in the null.

This electric dipole effect can be canceled by shielding the loop and arranging the shield so that the currents induced in each vertical member of the shield do not appear across the input impedance of the receiver. This arrangement, shown in Figure 10-16, can be compared to the simple balun (or balance-to-balance conversion), the balanced loop output becoming unbalanced, and the unbalanced electric dipole effect induced on the shield, canceling in the loop.

10.3.2 Balanced Loops

The electric dipole effect, which is also known as "antenna effect," "electrostatic error," "vertical component," and "stray vertical," can also be
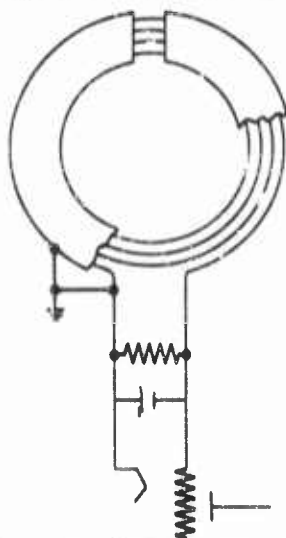
Fig. 10-16. Shielded loop for reduction of dipole effect.
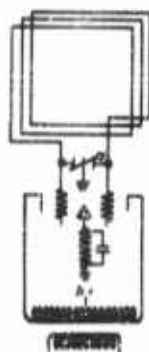
Fig. 10-17. Balanced rf amplifier for reduction of dipole effect.
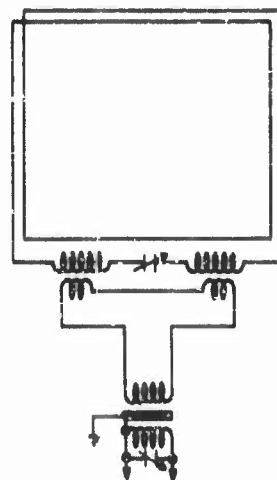
Fig. 10-18. Shielded transformer for reduction of dipole effect.

reduced by various balancing schemes, some of which are shown in Figures 10-17 and 10-18.

10.3.3 Pancake Loops

For a multiturn loop, each side consists of a number of conductors, each turn of which occupies a successively more remote position from a signal source, broadside to the loop. The difference in position produces a differential phase between turns so that a null is no longer present for this position.

This effect can be reduced by the use of the pancake loop shown in Figure 10-19. The pancake loop, however, has more pickup of the electric dipole mentioned above; hence, greater care must be directed toward the reduction of this associated difficulty.

10.3.4 Loop Systems

To avoid the mechanical rotation of a loop, the systems proposed by Bellini and Tosi utilizing an inductive goniometer to sequentially sample a pair of inclined antennas (as shown in Figures 10-3 and 10-4) have been applied to crossed loops. Figure 10-20 shows such a crossed-loop array as used in the Navy DAK equipment. The DAK operates in the frequency

FIG. 10-20. Crossed - loop direction finder (DA K) for 250-1500-kc range.
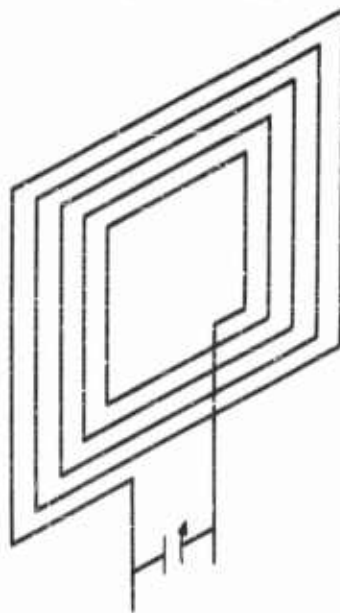
FIG. 10-19. Pancake-loop antenna to reduce quadrature pickup.

range 250 to 1500 kc in two bands split at 610 kc. The collector system is mechanically arranged to mount on a ship's superstructure. The bearing indication is a cathode-ray indicator which shows a propeller pattern of the type discussed in Section 10.8.

In addition to the visual indication, it is possible to mechanically rotate the goniometer to produce an aural null.

The goniometrically rotated patterns have a number of advantages over the mechanically rotatable loop. The rotatable loop, experience has shown, requires about 10 seconds to obtain a bearing on a signal having about a 10-db signal-to-noise ratio. For a manually rotated goniometer this time was reduced to about 3 seconds, while with a cathode-ray indicator, only about ½ second was required. These are approximate times since much depends on the observer's skill, on propagation conditions, and on the signal characteristics.

A crossed-loop direction finder using a goniometer and operating in the hf range (1.5-22 Mc) is shown in Figure 10-21. This is a Navy DAU type and is arranged for mounting atop stub mast or aftermast of a vessel. The bearing presentation is a propeller pattern displayed on a cathode-ray scope. The goniometer design and indicator details are presented in Sections 10.7 and 10.8.

Loop antennas have also been used at the higher frequencies. Figure 10-22 shows a loop antenna used on CXGJ, a shipboard direction finder operating in the frequency range of 30-100 Mc.

This equipment uses a shielded rotating loop (Reference 11) inside a streamlined plastic housing. The display is a cathode-ray tube presentation with sense being obtained by adding the output of a dipole to the loop signal.

10.3.5 Loop Arrays

Loops may be used as elements in an array which can then be rotated to determine the direction of arrival. Loops can be grouped as linear multi-element broadside, binomial, Dolph-Chebyshev arrays, and can be stacked vertically for more directivity in a vertical plane.

The binomial array may be used to reduce the minor lobe difficulty. In this type, all the radiators are also connected in phase, but the center elements of the array contribute more heavily to the input of the receiver.

The design data for these and also the Dolph-Chebyshev arrays may be found in a number of references (e.g., page 692 et seq. of Reference 2).

10.3.6 Spaced and Opposed Loops

To reduce the error due to horizontal polarization, many arrangements of two or more loops have been suggested. One of these systems used by three
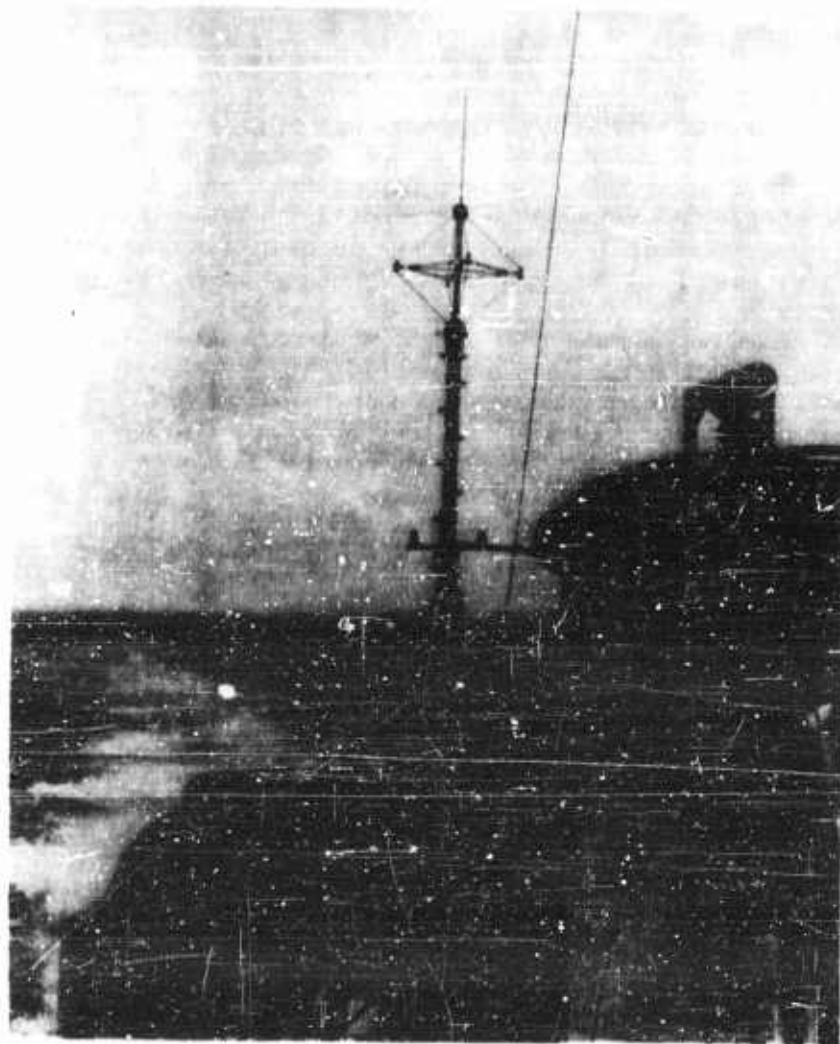
Fig. 10-21. Crossed-loop direction finder, DAU type, 1.5-22.0-Mc range.

FIG. 10-22. Rotating-loop direction finder for airborne use (CXGJ), 30-100-Mc frequency range.
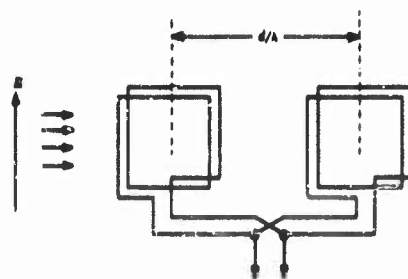
FIG. 10-23. Spaced and opposed loops for reduction of polarization error.

Independent investigators (Franklin, Eckersley, and Weagant) is shown in Figure 10-23. The two loops are arranged in coplanar fashion and connected out of phase. A signal with vertical polarization arriving in the plane of the loops produces an output dependent on the spacing of the loops; in fact, such spacing introduces a differential phase so that the combined outputs of both loops are no longer out of phase. The horizontal component, which, arriving from a direction broadside to the array, ordinarily causes difficulty, is canceled in the combined out-of-phase outputs. In this manner, for the broadside arrival of a signal, minima exist for both polarizations. Spurious minima can still exist at angles of the broadside positions since there will be positions where the vertical and horizontal polarizations may cancel, producing an extraneous minimum.

10.3.7 Dipole Types

An electric dipole, arranged horizontally, will have a figure-eight pattern and may be used as a DF antenna. Since there is a phase change of 180° as the dipole is rotated, the output can be combined with a horizontal loop to

obtain a sense voltage and produce a cardioid to avoid an ambiguity in direction of arrival. The pattern for such a dipole is shown in Figure 10-24.

Fig. 10-24. Directional pattern of an electric dipole arranged horizontally. (a) Dipole. (b) Cross section of directional pattern. (c) Perspective view of directional pattern.

The usual precautions must be taken to assure a balanced output, since an insufficient or asymmetrical counterpoise, or irregular nearby reflecting objects, will not only distort the pattern but also can cause pickup of vertical polarization generally by causing unbalance and pickup by the transmission lines. Poor connectors, faulty cables, or improper baluns, can all produce pattern distortions by what may be considered as analogous to the "electric dipole effect," "antenna effect," and "electrostatic error" for the previously discussed case of the loop.

10.3.8 Rotating Cardioid

If two antenna elements are arranged at a distance between $\lambda/8$ and $\lambda/4$ at the operating frequency and are adjusted in phase, an approximate cardioid can be obtained.

This pattern can then be rotated to produce at the output of a receiver what is essentially a sine wave with a frequency equal to the antenna rotation rate. The phase of the sine wave thus obtained is compared with the position of the loop by one of many methods to determine the direction of arrival.

Instead of deriving a voltage from each element directly, a parasitic element may be used in place of one antenna. The parasitic element may then be rotated about the active element as a center, therefore requiring no rotating feed and thus simplifying the mechanical and electrical structure.

A rotating cardioid antenna for direction finding in the frequency range 225-400 Mc is shown in Figure 10-25. This is one of two similar antennas used in a space-diversity arrangement in the AN/CRD-6 equipment.

Each antenna consists of a stationary vertical dipole around which is rotated, at 1800 rpm, a parasitic reflector. The dipole is 19 inches long; the reflector is 22 inches long and located $3\frac{1}{2}$ inches from the dipole. As the reflector rotates, it amplitude-modulates the received signal at the rotational rate. Coupled to the drive motor M is a two-phase generator G with output frequency equal to the amplitude-modulating frequency.
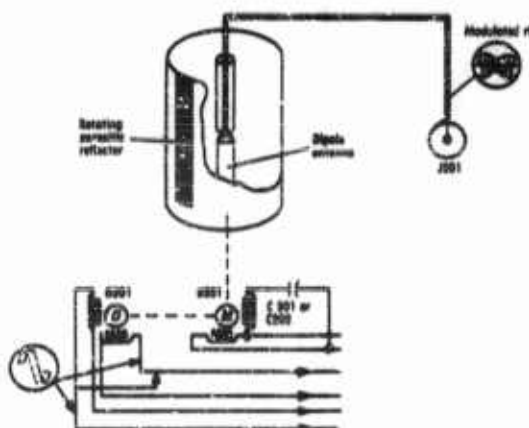
Fig. 10-25.   Cardioid antenna as used on CRD-6 uhf direction finder.

After detection, the phase of the antenna pattern modulation is compared with the phase of the reference frequency produced by the generator. The phase difference between the two produces an indication on the azimuth indicator.

Figure 10-26 shows a simplified block diagram of the system arranged as a two-antenna space-diversity system. The comparator CM-23/gr samples
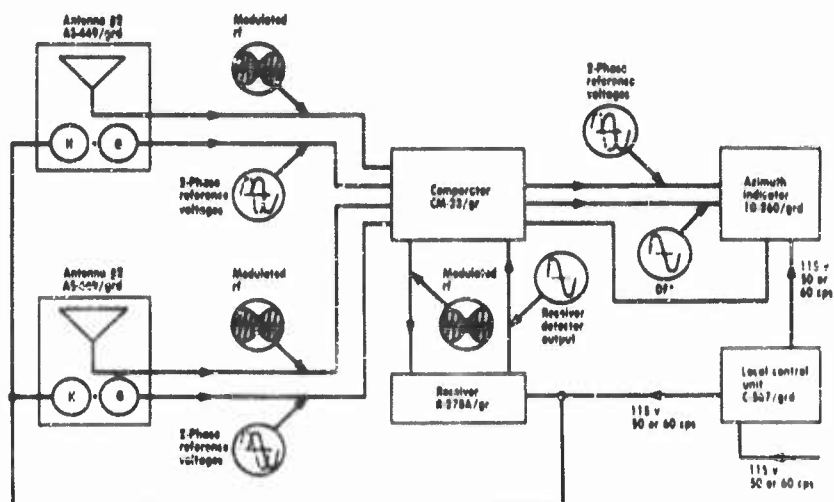
Fig. 10-26.   Functional block diagram of direction finder AN/CRD-6. DF* is the receiver output during the operating period only.

each antenna, determines which signal is the stronger, and connects this stronger signal to the azimuth indicator.

After the antenna with the stronger signal is selected, the comparator continues to sample each antenna and operates to keep or exchange the connected antenna depending on which of the sampled outputs is the larger.

Figure 10-27 is a simplified block diagram of the azimuth indicator which
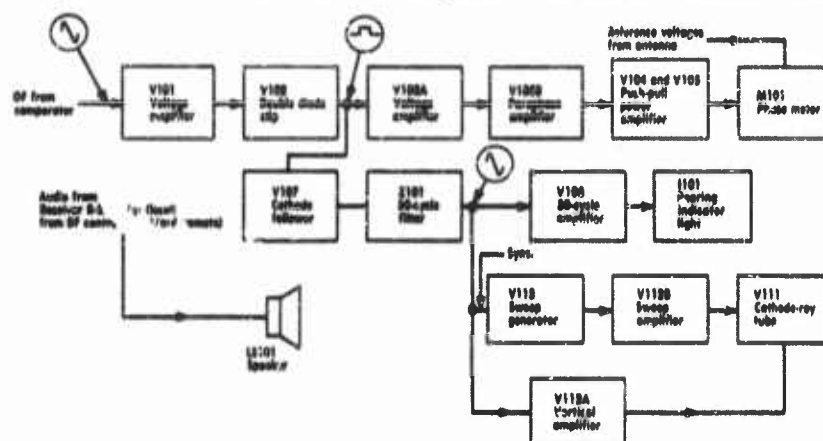
Fig. 10-27. Azimuth indicator for AN/CRD-6 block diagram.

feeds three visual indicating devices and one aural unit (speaker LS101).

The phase meter M101 is a mechanical device which indicates the bearing visually. The signal can also be examined visually by the cathode-ray oscilloscope V111. A third visual aid, I101, is a bearing indicator light which warns the observer when the comparator is sampling and a reading may not be taken.

10.3.9 Irregular Pattern Type

To illustrate that the pattern need not be regular, symmetrical, and simple, a DF apparatus with a wide aperture and having as complex and irregular a pattern as possible is shown in Figure 10-28.

A receiver R_1 is connected to antenna #1 and so arranged that as its oscillator is tuned to f_1, the desired signal, a remote oscillator generates a signal at f_2 on an adjacent band to f_1. This calibrating frequency f_2 is transmitted toward the DF antenna #1.

The receiver R_1 is arranged to have two i-f strips at $f_{osc} - f_1$ and $f_{osc} - f_2$. The output of each strip is detected in a separate detector. The output of one detector, in this case, that containing the demodulated f_2, is applied to a tape recorder with a movable pickup head and so arranged to

Fig. 10-28. A correlation direction finder using an irregular pattern.
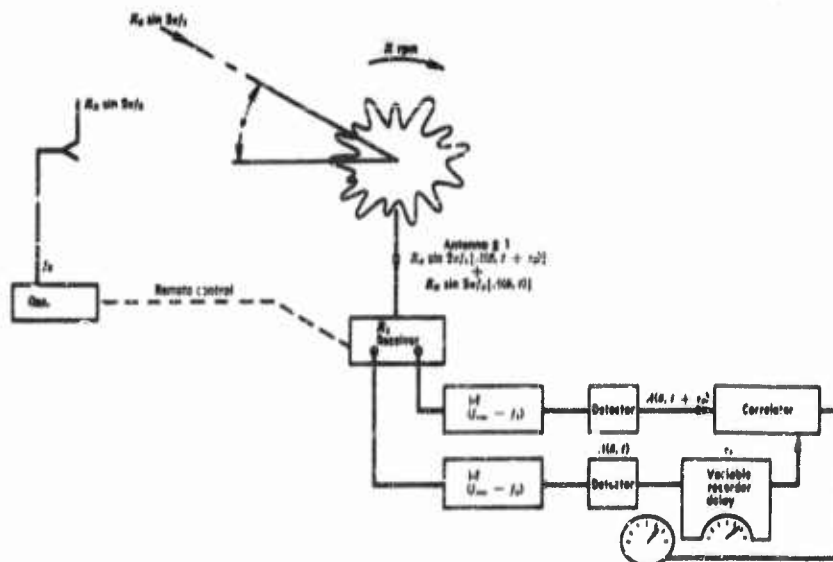
introduce a delay $t = 1/R$ from t_0 to $t_0 + (1/R)$ minutes, where R is the rotational rate of the antenna per minute, and t_0 is dictated by the mechanical design of the recorder.

The delayed output of the recorder is fed to one input of a correlator, and the demodulated f_1 signal is fed to the other input.

The correlator multiplies both inputs together and smooths, performing the operation

$$\phi_{f1/2} = \frac{1}{2T} \int_{-T}^{+T} K_1 A(\theta_1 t + \tau_\theta) \cdot K_2 A(\theta_1 t + \tau_v)\, dt$$

The variable delay τ_v introduced by the recorder is adjusted to maximize $\phi_{f1/2}$. This maximum occurs at $\tau_\theta = \tau_v$, and this delay is transformed to a bearing of the desired signal relative to the bearing of the calibrating oscillator f_2 by the simple relation

$$\theta^\circ = (\tau_v/R) \cdot 360^\circ$$

where, as above, R is the rotational rate of antenna #1 per minute.

10.8.10 Adcock Systems

In 1919 F. Adcock patented a DF antenna array in which two orthogonal figure-eight patterns were obtained by two pairs of spaced vertical antenna elements utilizing a goniometer similar to the Bellini-Tosi system and described in Section 10.7.

The goniometer was arranged so that varying portions of the output of each pair were sequentially sampled. In this manner, a rotating figure-eight pattern was obtained. The 180° ambiguity was resolved by adding a sense voltage from a center antenna.

With an Adcock system, the error due to extraneous horizontal polarization present can be minimized to a few degrees; however, a number of precautions are necessary. The symmetry of the electric field of each antenna must be preserved, the ground plane must be smooth with uniform conductivity, and the shielding or balancing of all horizontal cables and feeds is of great importance. The most important of these is the last, namely, the shielding or balancing of the horizontal cables.

Several variations of the Adcock system are in use. At frequencies above 100 Mc, it is practical to use dipoles with a balanced feed from each element as in the URD-4, Figure 10-29(a). This array is sometimes called an elevated "H."
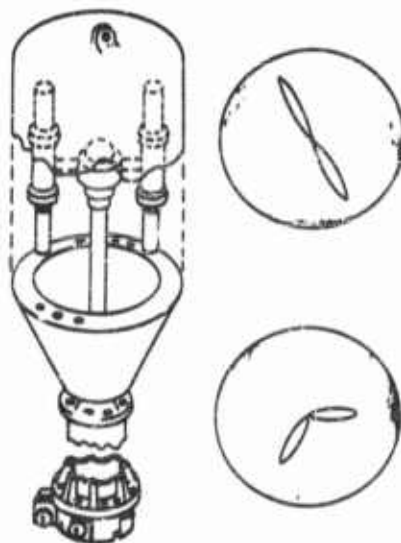
Fig. 10-29. (a) AN/URD-4. (b) Bearing pattern, no modulation. (c) Sense pattern, no modulation.

The URD-4 operates in the frequency range 225.0-399.9 Mc. It is a mechanically rotated pair of dipoles motor driven at 825 rpm. A two-phase generator similar to the arrangement previously described with respect to the CRD-6 (Figure 10-25) is used to generate two voltages in quadrature, that is, sine and cosine functions, for application to the plates of a cathode-ray tube to form a circular sweep pattern.

The 180° ambiguity of the symmetrical pattern is resolved by outphasing one of the elements of the antenna with respect to the other and shifting the resulting bent-back figure-eight pattern through 90° on the CRT so as to point toward the correct azimuth. In Figure 10-29(b) is shown a CRT pattern on a strong unmodulated signal and Figure 10-29(c) illustrates the sense-resolving feature.

Two "H" pairs can be arranged in a crossed configuration to form an elevated "H" Adcock array. The pairs can be sampled by a goniometer, either motor driven or manually operated. The sequential sampling method can also be replaced by a simultaneous sampling of both pairs, but systems of this sort are discussed under instantaneous types (Section 10.6).

When Adcock antenna arrays are used, with either a sequential or simultaneous sampling of the output of each of two crossed pairs, the precision with which the ideal figure-eight patterns are maintained will determine the ultimate accuracy of the DF equipment.

If the elements of a pair are arranged close to each other, that is, less than 0.1 λ apart, then the antenna pattern is substantially sinusoidal. A goniometer designed to operate from two crossed pairs having sine and cosine patterns with respect to a reference azimuth will, under these conditions, show a negligible error.

The close spacing, however, will result in an extremely small effective aperture and therefore the sensitivity of the system will be very low, since the aperture is substantially proportional to d/λ when d/λ is small ($d =$ spacing).

In a four-element Adcock for a maximum error of $2°$ at the highest frequency to be received, the spacing between elements of a pair should not exceed 0.28 λ.

This error will increase with downcoming signals, since the effective spacing varies as sec ϕ, where ϕ is the elevation angle of the arriving signal. The "spacing" error increases with frequency and therefore imposes an upper limit in frequency range beyond which the error becomes excessive.

In order to reduce this error, eight-element Adcock arrays have been developed. These arrays have four pairs of antenna elements symmetrically disposed on a circumference. Two methods are used to connect these antenna pairs.

In the first method, each odd-numbered antenna is connected to an adjacent antenna in the same direction. Each interconnected adjacent pair forms an Adcock element being connected out of phase with a diametrically opposite paralleled pair.

The second method uses a polyphase goniometer, the details of which will be shown in a later section.

An eight-element antenna system using the polyphase goniometer is shown in Figure 10-30. This equipment is known as the GRD-5 and covers the vhf band.

The eight-element Adcocks have a reduced "spacing" error since the output of each pair is used over a restricted angular coverage. Since the distortion of the figure-eight diagram is least in the vicinity of its maximum and minimum values, the error is minimized by this restricted use of each pair.

Figure 10-31 shows the spacing errors of multielement Adcock systems using polyphase goniometers (Reference 12). In this figure

N = total number of elements

n = number of pairs

d = spacing between elements of pair

λ = wavelength to be received

$\triangle$ = maximum "spacing" error in degrees

Figure 10-32 shows the antenna system of the SCR-291A. An equipment similar to this, the DAJ, was an important factor in nullifying the German submarine threat during World War II. This antenna equipment consists of two directional pairs and a sense antenna. Four antenna elements are located at the corners of a 25-foot square and the fifth at the center. Each element consists of a vertical wire supported within a 24-foot telescoping mast of plywood tubing. The antenna pairs are sampled by a motor-driven inductive goniometer and displayed on a 5-inch CRT indicator. The equipment covers the frequency range 2.0-10 Mc in three bands. The equipment has a $\pm 2°$ bearing readability for about $10\mu v/m$ field strength.

Fig. 10-30. An eight-element Adcock array for vhf use.

The display, which is of a propeller pattern type, is considered in Section 10.6.1.

At the hf range for permanent installations, the most practical arrangements are those in which the feeds are buried below the surface of the ground so as to reduce the standard wave error. The standard wave error is a measure of the immunity of the system to polarization error.

The *standard wave error* is defined as the error produced by a downcoming wave at an elevation angle of 45° and having equal electric field components which are, respectively, in and at right angles to the vertical plane which includes the DF antenna and the test transmitter, the phase relation of the two electric-field components being adjusted to produce maximum DF error (Reference 10).

The buried Adcock can further be improved by the use of buried wires in a counterpoise arrangement (Reference 13) as shown in Figure 10-33. Figure 10-34 shows the reductions in standard wave error for a transportable Adcock brought about by the use of radials of various lengths in the configuration of Figure 10-33. A loop transmitter was elevated 50 feet above the ground at a distance of 400 feet from the DF array. The soil conduc-
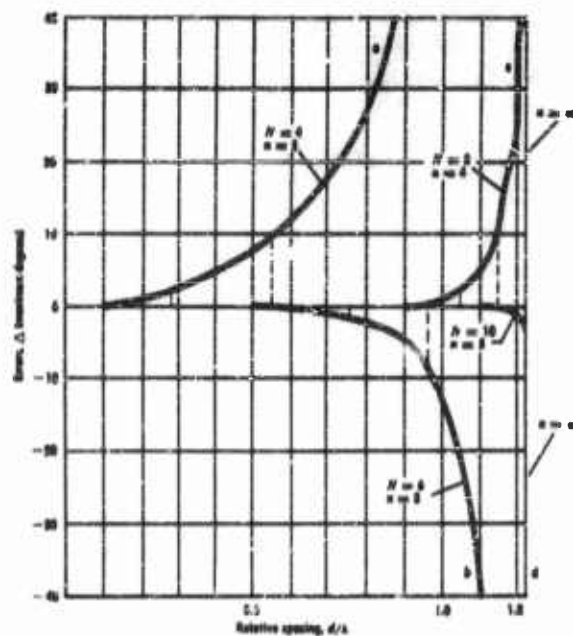
Fig. 10-31. Spacing errors of multiaerial Adcock systems with polyphase goniometer.
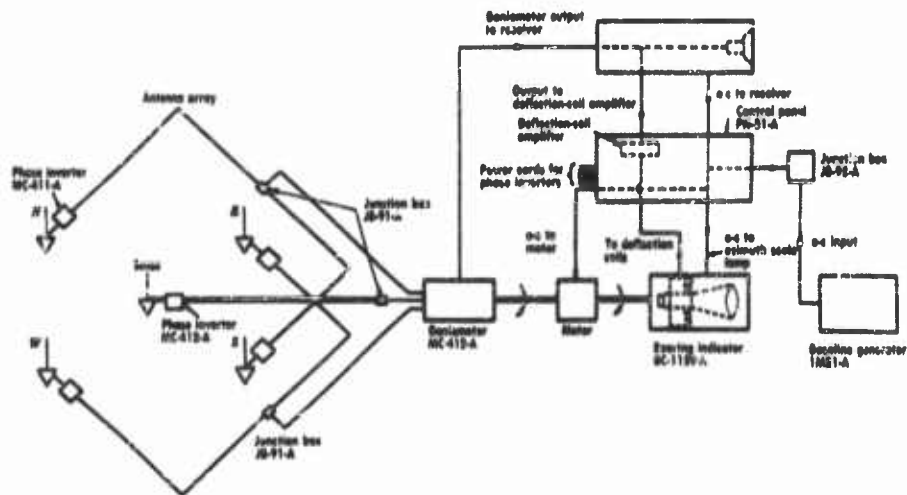
Fig. 10-32. Antenna system SCR-291A bf Adcock direction finder with block diagram of associated auxiliaries.
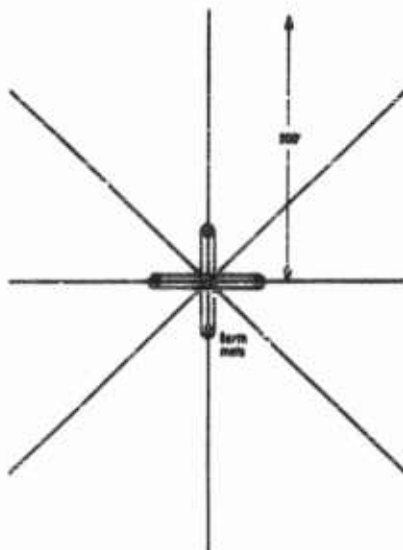
Fig. 10-33. Buried U antennas with counterpoise wires and earth mats for reduction of polarization errors.
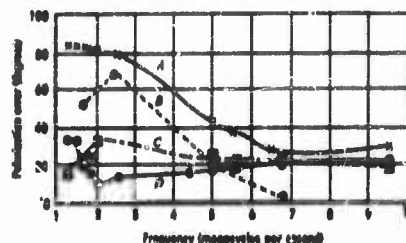
Fig. 10-34. Variation in polarization error with type of radial system, for DF station transportable Adcock A = with earth mats, no radials. B = with earth mats and four 52-foot radials. C = with earth mats and four 200-foot radials. D = with earth mats and eight 200-foot radials.

tivity was measured at ~ 10^{-8} mhos/m and the permittivity as 8.85×10^{-11} F/m at 5 Mc.

· From tests of this sort, it is found that substantial improvements are possible (60°-70°) by various aids when the site conditions are poor. However, if a good site is available, many precautions are unnecessary.

More will be said about counterpoise arrangements in Section 10.9.

10.3.11 Reflectors

To increase the directivity of a single antenna element such as a vertical dipole and not complicate the feed and phasing problem, a reflector is often used. When two flat sheets intersecting at an angle are used, a sharper pattern than a flat reflector is obtained. Figure 10-35 shows the field pattern for a square corner reflector with an antenna-to-corner spacing of 1.5 λ.

The rotating dish or reflector is employed in such equipments as the AN/APA-17 and AN/ALA-6. The rotating dish, in conjunction with a receiver, detector, and display unit, is utilized to provide a signal strength versus azimuth plot of the electromagnetic environment. The indicator provides a PPI presentation from which the angle of arrival is determined. The rotating dish principle is illustrated in Figure 10-36, where an ideal signal-strength plot is depicted.
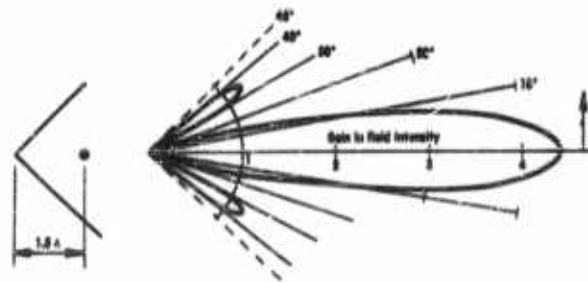
Fig. 10-35.   Calculated pattern of square corner-reflector antenna with antenna-to-corner spacing of 1.5 λ relative to λ/2 dipole in free space.
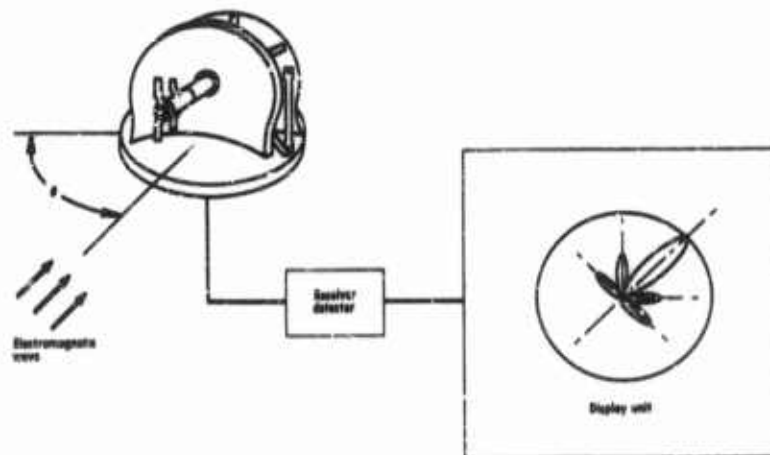
Fig. 10-36.   Rotating-dish direction-finding principle.

The antenna configuration utilized in the rotating-dish direction finder is shown in Figures 10-37(a) and 10-37(b). Both horizontally and vertically polarized antennas are alternately switched to the receiver to permit reception of both polarizations. Since the directional characteristics are in opposite directions, a 180° bearing shift is produced when switching between antennas. Compensation can be easily provided by reversing the indicator as one antenna is switched.

The rotating-dish direction-finding technique has been utilized for many years; however, it has a number of difficulties.

Since the signal-strength plot is one corresponding to the antenna pattern
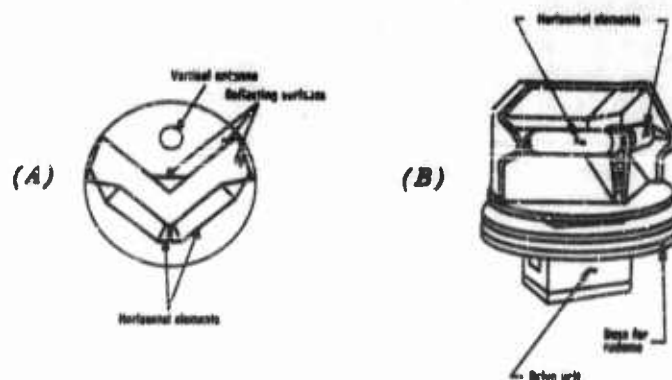
Fig. 10-37. Antenna configuration utilized in the rotating dish direction finder.

of the direction-finding antenna, patterns will vary as a function of frequency (the antenna is generally broadband).

If a rotating emitter (such as a radar) is intercepted, distortions in the signal-strength plot are quite probably due to the transmitter lobe structure which in turn gives erroneous azimuth data.

At frequencies below 1000 Mc, the dish becomes too large to permit its use on high-speed aircraft.

10.3.12 Miscellaneous Elements

Often it is advantageous to use antenna elements other than loops or dipoles.

Dielectric rods, helixes, terminated wave antennas, rhombics, half rhombics over an image plane, slots of many configurations, open-ended waveguides, dielectric-loaded and ridge-loaded horns, all have been used and offer certain advantages depending upon the application. The characteristics of such radiating elements are discussed in Chapter 29, "Antennas and Transmission Lines."

10.3.13 Wide-Aperture Systems—Sequential Amplitude-Measurement Types

As the radio wave propagates in space over an irregular terrain, it is absorbed, reflected, diffracted, and refracted. Irregularities along the ground, both natural and man made, distort the wavefront so that, at a given point, a measurement of the apparent direction of arrival may have large errors.

Figure 10-38 shows a representative condition when the field from a signal source R is affected by a strong reradiating source at S. Direction finders located at phase discontinuities A, B, C, D, and E would show large errors.
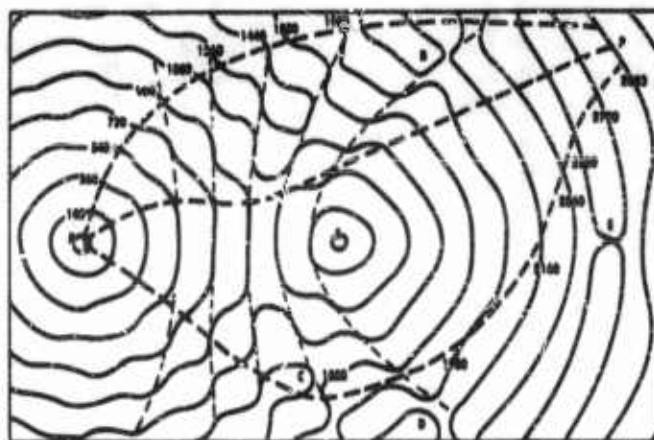
FIG. 10-38. Equiphase contours.

However, if a large aperture DF array is used extending well beyond the points of maximum phase disturbance, the error may be reduced substantially by averaging or by more sophisticated statistical procedures.

The problem of exploring the field over a large area, i.e., many wavelengths, is relatively simple at frequencies above 1000 Mc, since antenna directivity can easily be achieved by the use of arrays and parabolic reflectors; and this directivity can be used to separate the primary source from secondary or extraneous sources.

At the lower frequencies and in particular at hf, high directivity requires large physical antenna structures which cannot readily be rotated mechanically to determine the azimuth of the strongest signal.

Antenna arrays are discussed briefly in the foregoing text. The effect of the ground plane on the design will depend on the polarization being used. At the vhf and uhf bands, the ground can be considered as a perfect conductor for horizontal polarization. The reflected wave is under those conditions equal to the incident wave but phase reversed so that an out-of-phase image is produced.

For vertical polarization at grazing incidence, the reflection coefficient is essentially the same for both polarizations. For other angles, however, the phase and amplitude of the reflected wave changes and is also dependent on the electromagnetic properties of the ground.

10.3.14 Lobe-Switching Arrays

By proper phase-change switching, the major lobe of an array may be

made to move through an equal angle on either side of the plane normal to
the array. This lobe switching is so arranged that the amplitude of a signal
normal to the array remains constant since the antenna gain for this position
is constant for both lobe positions. The array is then rotated slowly while
the lobe switching is done rapidly enough so as to obtain several samples for
each incremental position. The rotational rate is therefore determined by
beam sharpness. Figure 10-39 shows the antenna pattern in both positions.

The advantage of a lobe-switched array lies in
the feature that the cross-over point or point of
equal amplitude in both lobe positions can be
selected at a high signal level and yet be a point
where the antenna pattern is changing most
rapidly with azimuth. The pattern will, however,
change with frequency so that, in general, lobe-
switching methods are limited to narrow frequency
ranges.

FIG. 10-39. Switched lobe di-
rection finder.

10.3.15 Beam-Shifting Arrays

At 1000 Mc and above, sharp beams can be
obtained from reflectors that are sufficiently small
(approximately 12 feet diameter for a 5° beamwidth). These dishes are still
small enough to be mechanically rotated for ship and shore installations, and
useful when the signal last sufficiently long to be intercepted by the beam
rotation. Systems which obtain the direction of arrival on a single pulse are
described in Section 10.6.

At frequencies below 100 Mc and in particular at hf (3-30 Mc), the
arrays become large and are only feasible for ground-based installations.

One of the early applications of beam-shifting arrays was the MUSA
system (Reference 14) used for the improvement of transoceanic communi-
cations in the presence of multipath and other propagation difficulties.

This system, called MUSA (Multiple Unit Steerable Antenna), used, in
its first phase, an endfire array of rhombics so arranged as to be steerable
in the vertical plane. A variable phase shifter was associated with each
antenna operating at the i-f of the receiver. Six rhombic antennas were used,
each antenna going to a mixer and a phase shifter. The outputs of the five
phase shifters (the first antenna not requiring a phase shifter since it served
as a reference phase) were then fed into a supplementary receiver which
provided the gain and second-detector functions. In this system several
branches of ganged phase shifters were provided so that several beams could
be synthesized at one time and several vertical angles could be viewed
simultaneously.

The MUSA system was extended subsequently to the generation of beams which could be steered in azimuth (Reference 15). This modification was used to study lateral deviations from the London-New York great-circle path and showed that, during "all-daylight" path conditions, multipath propagation was bunched in or near the great-circle plane. During periods of dark or partially illuminated path conditions, the propagation was not limited to the general vicinity of the great-circle plane but would, under conditions of ionospheric disturbances, even of moderate intensity, involve paths south of the great circle. At times, wide southerly deviations up to 75° were found from 5 to 15 Mc. Simultaneous deviations of about 30° were found at times on 9, 7, and 5 Mc.

Figure 10-40(a) shows the broadside array used for the formation of the horizontally steerable beams. The asymmetry which results due to the mutual coupling of adjacent antennas not existing at the end elements is partially compensated for by the use of dummy elements at each end.

Figure 10-40(b) is a schematic of the MUSA receiving apparatus; the output of each antenna is combined linearly at i-f. If further reduction of minor lobes were required, the combination of the outputs could be weighted in favor of the middle elements to approximate a binomial distribution.

The cage broadside MUSA was designed to have high directivity consistent with a single-lobed response. The highest frequency for which single-lobed response is required determines the spacing of adjacent elements which must be $\lambda/2$ at f_{max}. Figure 10-41 shows the calculated patterns at 10 Mc (f_{max}) and 5 Mc ($f_{max}/2$). The back lobes can be reduced if necessary by using unidirectional configurations for each antenna element.

If back radiation from each element is suppressed and the number of active elements is increased to eight, the patterns of Figures 10-42(a) and 42(b) can be obtained. These patterns were calculated for a synthetic array in which the phase at each antenna is recorded with respect to a reference antenna, and a computer used to synthesize, after the reception, the optimum pattern (References 16 and 17).

10.3.16 Circular Arrays

The use of DF arrays having narrow beams has a number of advantages. In the hf range, multipath signals can sometimes be resolved and a more accurate bearing obtained. In addition, the usual advantages of a large aperture array in giving good bearings under poor site conditions prevail. Whether a large aperture using doppler or a sharp beam is best depends on the distribution in azimuth of the reflecting objects, and on whether the total reflected energy comes from a few large objects or many small ones.

Returning to the wide-aperture narrow-beam type: the number of elements
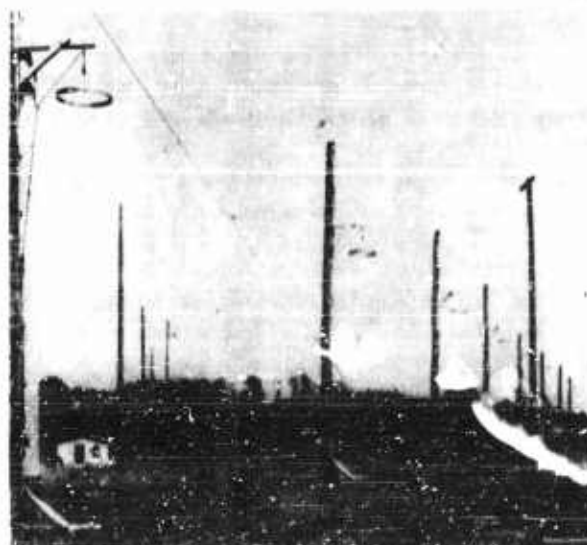
(A)
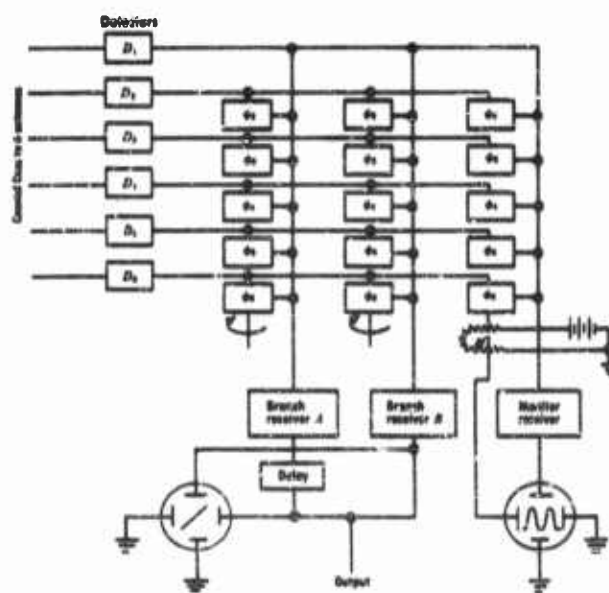
(B)

Fig. 10-40.   (a) Broadside array used in the MUSA system for horizon azimuthal
selectivity. (b) Schematic of MUSA receiving apparatus.
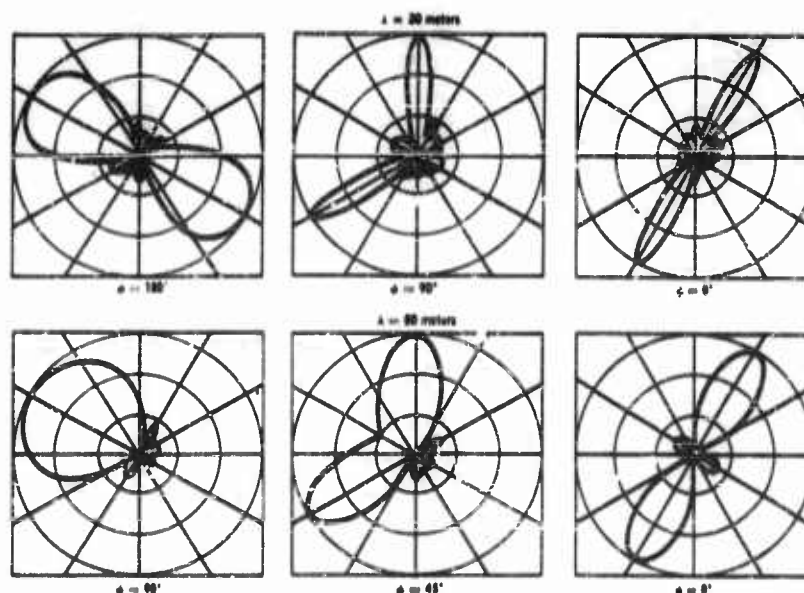
Fig. 10-41. Horizontal plane directional patterns of cage broadside MUSA. The image lobe is a result of the omnidirectional azimuth pattern for each cage element. ϕ is the fundamental variable phase shift: ϕ_2 of Figure 10-40(a).

In the MUSA can be extended and arranged in a circle in the event that a full 360° coverage is required. This is what the Germans did during World War II in developing the Wullenweber.

Instead of using lumped phase-shifting networks operating at i-f, the Wullenweber, at some sacrifice in sensitivity, used a goniometer with delay lines connected to about one-third of the elements, contiguously selected. The delay lines were adjusted to bring into phase the outputs of these adjacent antennas. The goniometer progressively selected, in a rotating fashion, each group of antennas. Figure 10-43(a) shows the basic principle of the Wullenweber. Figure 10-43(b) shows a 120-element array with a diameter of 1000 feet installed on the facilities of the University of Illinois near Urbana, Illinois.

With the Wullenweber, only a portion of the total aperture is used at any one time, usually approximately one-third of the elements. If more than one-third of the elements are used, the goniometer becomes complex, the longest delay lines become bulky and lossy, and the performance is not appreciably improved.

The goniometer and delay-line configuration present some serious design problems. Each antenna feeds the receiver through a transmission line which
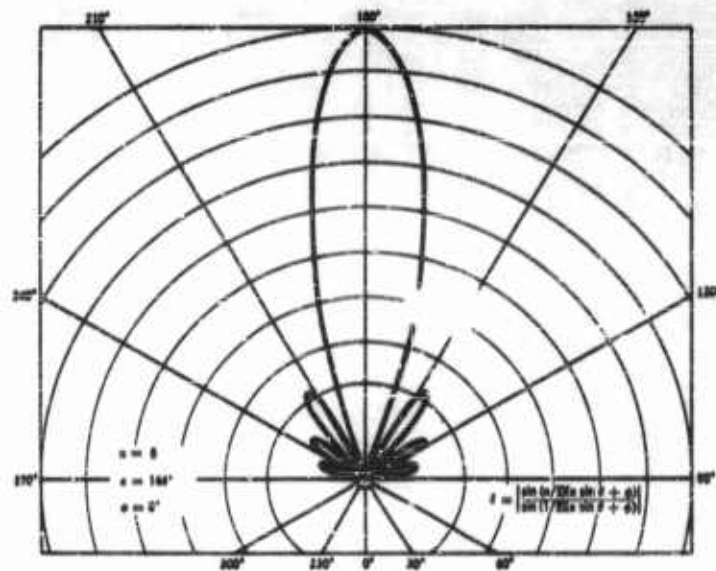
(A)
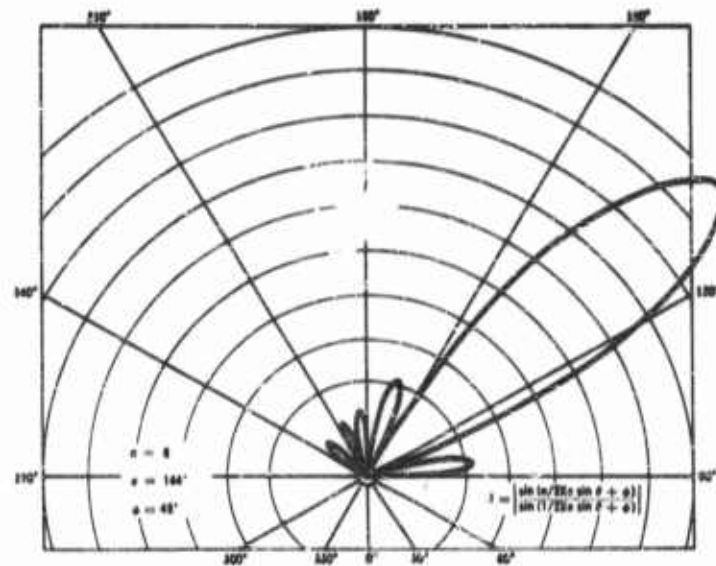
(B)

FIG. 10-42. Plots of receiver voltage vs. phase shift for sythesized eight-element section. $n = 8$; $\alpha = 144°$; $\theta = 45°$.
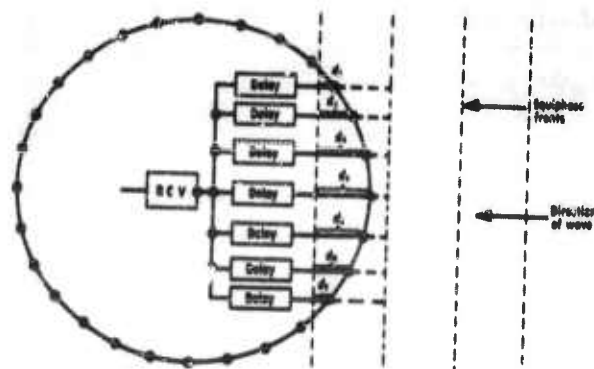
(A)

(B)

FIG. 10-43. (a) Basic principle of Wullenweber. (b) Wullenweber hf array with 120 elements located on the University of Illinois antenna test site.

introduces the required delay. The impedance which each antenna presents to the receiver junction is a function of the line length and the frequency being received. In order to prevent changes in impedance with frequency at the receiver input, each antenna is generally lightly coupled to its delay line, thus reducing the sensitivity of the system. If isolation amplifiers are provided at each antenna, then such amplifiers must maintain a constant or calibrated phase characteristic to an extreme accuracy over the frequency range.

The proper design of isolation amplifiers at each antenna would moreover permit the simultaneous use of the array for more than one signal, or the continuous monitoring of one signal source while other azimuths are being examined for new signals.

Instrumental accuracies of better than $\frac{1}{2}°$ can be obtained with 30-λ apertures above 100 Mc.

At hf, the limit in accuracy will generally be due to propagation phenomena. At ranges from 0 to 20,000 km, substantial deviations from the great-circle routes have been experienced (References 18, 19, and 20). It has been suggested (Reference 21) that this effect is a lateral deviation produced by tilts in the ionosphere caused by diffraction processes in the F_1 layer. The tilting of the reflecting layer is insufficient to produce the sizable deviations. These phenomena are illustrated in Figures 10-44(a), (b), and (c). Figure 10-44(a) shows the end-on view of the prism effect in the refracting layer, producing a lateral deviation to the left. Figure 10-44(b) shows the
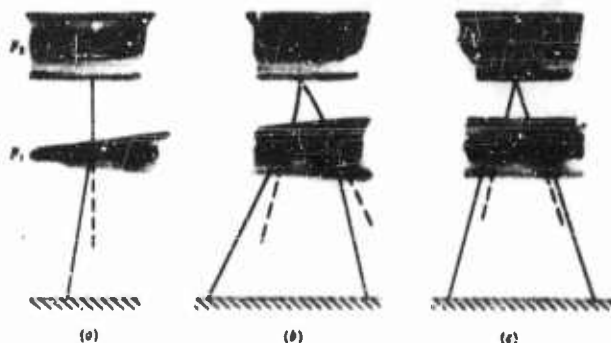
Fig. 10-44. (a) The end-on view of the prism effect in the refracting layer. (b) The side view of the refraction phenomenon in the F_1-layer with reflection from the F_2-layer. (c) The symmetrical refraction from a uniform F_1-layer.

side view of the refraction phenomenon in the F_1 with reflection from the F_2 layer. Figure 10-44(c) shows the symmetrical refraction from a uniform F_1 layer.

These effects are slow variations due to gradual changes in ionosphere configurations averaged over $\frac{1}{2}$ to 1 hour. There are, of course, the rapid fluctuations due to interference effects of multipath propagation. Figure 10-45 shows the results of tests with a wide-aperture interferometer on hf iono-sphere-reflected signals (Reference 22).

The question is still to be resolved whether a wide-aperture system of N elements located in one area is capable, at hf, of yielding better accuracies on distant signals than that number of elements arranged in a number of Adcocks with space diversity.
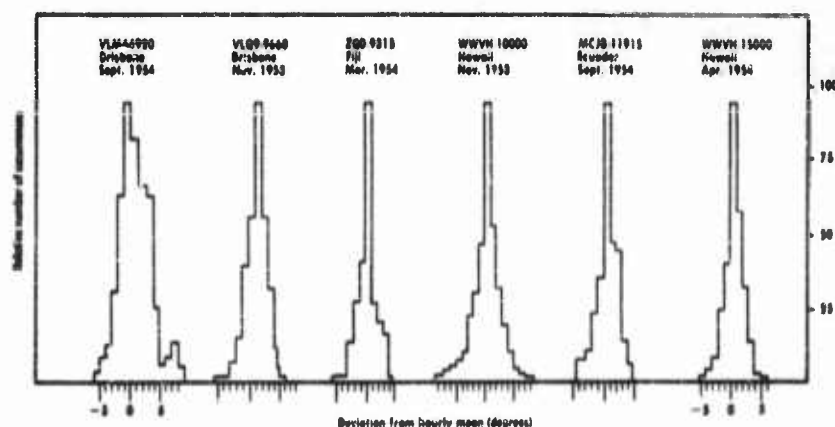
FIG. 10-45. Histograms of bearing deviations from the mean for each hour of the months shown. At the top is given the station, frequency in kilocycles, location, and month.

If the propagation is by ionosphere reflection and if the disturbances in the phase front of the electromagnetic field due to a plurality of reflecting areas are grouped closely together, then these phase-front disturbances change very slowly with distance in the vicinity of the receiver.

For example, if two images of the transmitter—one as a result of the F_2-layer and the other as a result of the F_2-layer but refracted by the F_1-layer—are considered, these rays may at a time and at a distance of 2000 km be separated by $5°$. Under these conditions the phase front at the receiver would be distorted by cyclical phase changes at distances of

$$\delta = \lambda/\sin 5°$$

For example, if $\lambda = 30$ meters, the quantity δ is

$$\delta = 30m/0.087 = 345 \text{ meters}$$

Attempts to average out the disturbance would require an aperture several times δ, depending on the residue error to be tolerated, or an aperture of over 1000 meters at $\lambda = 30$ meters.

If the reflecting or refracting objects are remote and close together, then a wide-aperture system would be impractically large. If, on the other hand, local reflecting objects or a poor ground plane is encountered and, in particular, if the azimuth of these reflecting objects is substantially different from that of the desired signals, then medium- and large-aperture systems will substantially reduce the error.

The design of a wide-aperture system should take into account the general direction in which the signals are expected to arrive and the locations of various unavoidable reflecting objects. Accurate data are required (Referen-

If the reflecting or refracting objects are remote and close together, then ces 23 and 24) on ionosphere statistics for daily and seasonal periods and under conditions of unusual magnetic and solar activity.

10.3.17 Doppler Direction-Finding Systems

The wide-aperture DF system using doppler techniques is considered under sequential phase-measurement systems (Section 10.4).

10.3.18 Multilobe and Interferometer Types

In airborne applications for high-speed aircraft, at frequencies below 200 Mc, it is difficult to design DF systems with reasonable accuracies ($\pm 5°$) and not introduce excessive aerodynamic drag. If two widely separated antennas are used to create a multilobe structure, the bearing of a signal can be obtained by observing the signal variations as the aircraft moves through the signal. This system is sometimes called the two-aerial interferometer. This technique is used on the ASQ-18 airborne reconnaissance system operating from 70 to 1000 Mc.

In this system, the antenna elements are spaced at a distance d, which is usefully 3 to 10 λ. It can be shown that the resulting interference pattern will contain nulls which occur at azimuth angles given by (see Figures 10-46 and 10-47):

$$\beta n = \sin^{-1} (n\lambda/d)$$

where βn = null angle

n = 0, 1, 2, 3, etc.

λ = wavelength

d = distance between elements
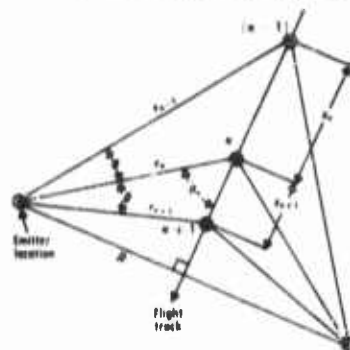
Fig. 10-46. Calculation of emitter location.

$$B_n = \tan^{-1} \frac{X_n + X_{n+1}}{X_r - X_{n+1}} \tan \theta$$

$$r_n = \frac{\sin (B_n - \theta)}{\sin \theta} X_n$$

The angle between successive nulls is

$$\theta = (\beta n + 1) - \beta n = \sin^{-1} \frac{(n + 1)\lambda}{d} - \sin^{-1} \frac{n\lambda}{d}$$

If the antenna spacing is large compared to one wavelength,

$$\theta = \lambda/d$$

Utilizing the known position of nulls in the azimuthal interference pattern of two or four antennas and correlating this with the navigation data and intercepted signal strength for a given emitter, it becomes possible to determine the direction of the emitter when it has gone through three successive nulls. If the location of the emitter is approximately known beforehand, then a directional cut can be obtained from one or two null recordings.

In practice the interference pattern generated by two elements will be modified by the individual radiation patterns of the elements. However, as long as these are essentially omnidirectional in regions of interest, and as long as unwanted nulls are minimized, the interferometer will yield much useful information.

Fig. 10-47. Antenna-radiation-pattern nulls as a function of frequency for wing-tip antennas on an RB-47E aircraft.

This method makes no pre-judgments of the antenna pattern of the emitter whose direction is sought, but requires only that the interferometer be moved in a reasonably straight line and that it be illuminated often enough to trace out the null modulation which it has established by its own construction.

10.3.19 Lens Types

While various types of lenses can be used to obtain a shaped beam, the Luneberg lens is the only one which has an important use in microwave direction finding.

R. K. Luneberg suggested (Reference 25) a nonhomogeneous lens for providing a 360° scan by moving only the pri-

mary feed. Such a lens must be radially symmetrical. The refractive index must then vary with the radial distance from the center to the circumference. The design of the lens (Figure 10-48) and the choice of the refractive

FIG. 10-48.   Ray paths and phase fronts in Luneberg lens.

index must be such that a plane phase front impingent on the lens circumference will focus the wave at a point F, which is the far end of a diameter perpendicular to the phase front.

To fulfill this condition, the integral of the refractive index times the element ds must be the same for all paths from the aperture AB (see Figure 10-48) to the feed F.

By providing a mutiplicity of probes or feeds which consist of dipoles arranged at an angle of 45° to the plane of the lens, the system can be made to operate instantaneously. The dipoles are arranged at 45° so as to be "transparent" for the wave impingent at that portion of the circumference. Figure 10-49 shows a spherical Luneberg lens similar to that used on the

FIG. 10-49.   Spherical Luneberg lens similar to that used on AN/DLD-1.

AN/DLD-1 microwave reconnaissance set, which is an airborne equipment operating in the frequency range 1000-40,000 Mc. Three Luneberg lenses are used to cover the range 1000-9100 Mc, while an array of horns covers the range 9100-40,000 Mc.

The Luneberg lens as used in the AN/DLD-1 equipment has fifteen outputs each connected to a video receiver. Its design is such as to obtain a bearing on a short-duration radar signal since it operates on an instantaneous comparison of adjacent sectors rather than on a sequential exploration or scanning process. The AN/DLD-1 is therefore classified as an instantaneous type and will be described in Section 10.6.

10.4 Sequential Phase-Measurement Types

The direction of arrival of a signal can be determined by moving an antenna from point to point in a horizontal plane and noting the phase change that results. A phase change of 2π radians per second is equivalent to a doppler frequency shift of 1 cycle. The field may be probed either by moving an antenna mechanically, or by having a fixed deployment of antennas and switching sequentially from one to the other.

Figure 10-50 shows the conditions for the mechanical movement of a probe in a circle about a vertical axis O.
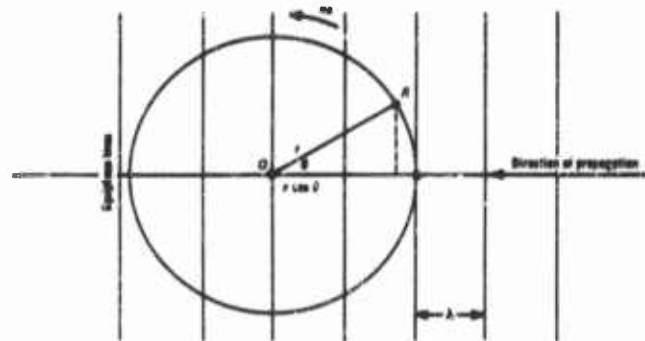
Fig. 10-50. Sequential phase-measurement (doppler) direction finder.

Let the field picked up by an antenna element in a stationary condition be

$$e = E_0 \sin \omega t = 2\pi f_o$$

where f_o = frequency of rf signal.

When the antenna element is rotated by an angle θ, this voltage is subject to a phase shift $(2\pi/\lambda) r \cos \theta$ from the origin and becomes

$$e = E_0 \sin \left(\omega t + \frac{2\pi}{\lambda} r \cos \theta \right) \tag{10-9}$$

where $\lambda = v/f_0$.

But if θ is rotated at a uniform rate ω_R and $\theta = \omega_R t$, then

$$e = E_0 \sin \left(\omega t + \frac{2\pi}{\lambda} r \cos \omega_R t \right) \tag{10-10}$$

If the rotating antenna output is applied to a linear discriminator or phase detector so that the output

$$e_D = K \cos \omega_R t$$

where K is dependent on the circuit constants, then

$$de_D/dt = 0$$

when $\theta = 0°, 180°, 360°$, etc.

Therefore, the direction of arrival can be determined by observing when the doppler shift is either zero (when the antenna is moving tangential to the equiphase lines) or when the doppler shift is a maximum (when the antenna is moving in the direction to or from the signal source).

In the case of continuous sampling of the output with a sinusoidal phase detector, one has

$$D_0 = \sin \left(\frac{2\pi}{\lambda} r \cos \omega_1 t \right) = \sin (R \cos \omega_1 t)$$

where the quantity $R = (2\pi/\lambda)r$ is half the aperture in electrical radians. The phase-detector output yields a series of odd harmonics, since, using a new reference time,

$$D_0 = \sin (R \sin \omega_1 t)$$
$$= 2 [J_1(R) \sin \omega_1 t + J_3(R) \sin 3\omega_1 t + J_5(R) \sin 5\omega_1 t + \ldots$$
$$+ J_{2n+1}(R) \sin (2n+1)\omega_1 t + \ldots]$$

Analysis of this is found in Reference 26. Analyses of other cases for the sequential examination of the output of a number of antennas arranged on the circumference of a circle are found in References 27, 28, 29, and 30.

When the sampling of the phase is by a step-by-step examination of the output of a series of antennas, the samplings may be considered as a square-

wave phase modulation of the signal; this process produces a number of sidebands, the spectrum being related to the geometry of the sampling.

To avoid distortions due to the phase-detector characteristic, it has been proposed to compress the phase changes by taking the difference in phase among several antennas (Reference 26).

A doppler system using a rotating sampling antenna moved over the circumference of a circle of many wavelengths will encounter a number of phase-front distortions. The distance between the phase perturbations is determined by the angle between the direct signal and the reflecting object, and the magnitude of the reflection.

As the field is explored, the apparent equiphase front will move back and forth through the true phase front. An averaging process will tend toward an error which approaches zero as the exploring path is made larger and larger. The residual error θ_R can be shown to be approximately

$$\theta_R = \frac{D}{\lambda} \tan^{-1} \frac{E_R}{E_D}$$

where $D =$ diameter of sampling circle in meters
        $\lambda =$ wavelength of signal in meters
        $E_R =$ field due to reflecting object
        $E_D =$ field due to direct signal

There are a number of direction finders available here and abroad utilizing the doppler principle. The principle of operation is simply to sequentially switch from element to element of the array and, after extracting the doppler modulation thus introduced, to compare this with a reference voltage synchronized with the antenna commutating rate. The phase of the doppler deviation with respect to the reference voltage indicates the direction of arrival.

The various systems differ chiefly in the use of a mechanical or electronic commutation, frequency range covered, and the type of indication.

The antenna arrays generally have from 12 to 30 dipoles less than $\lambda/2$ apart and spaced equally around the circumference of a circle. Figure 10-51 shows an antenna array of 30 dipoles operating in the uhf range.

Figure 10-52 shows a block diagram of a system using electronic commutation. This equipment uses a dual-channel, double-superheterodyne receiver with a common oscillator. One receiver section amplifies the sequentially switched DF antenna signal, while the other amplifies the auxiliary- or reference-antenna output. The second i-f outputs of the two channels are applied to a phase-comparison circuit, which extracts the bearing information and provides an input for the bearing display and indicator unit.

Fig. 10-31. Wide-base-line doppler direction finders for vhf range using 30 dipoles.
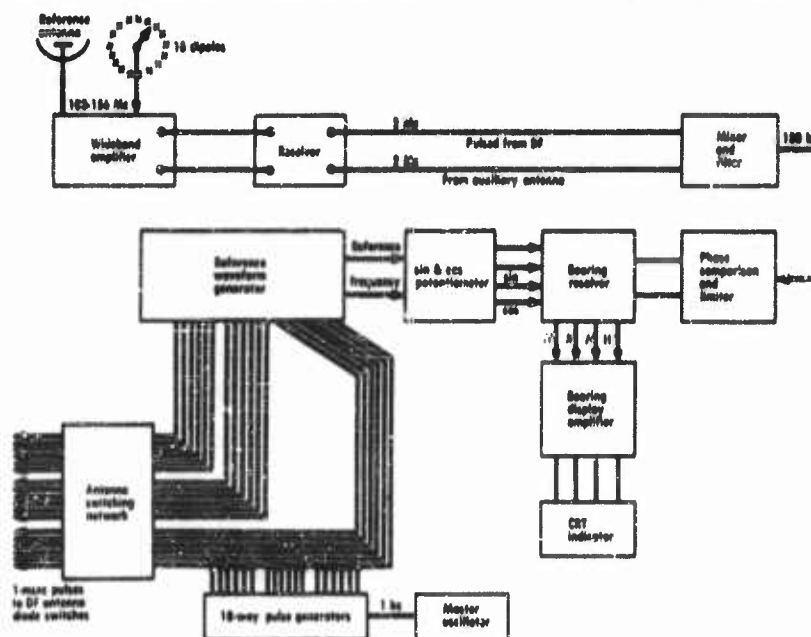
Fig. 10-52. Doppler direction finder with 18 antennas operating in the vhf band using diodes for commutation.

The phase-comparison circuit demodulates the DF phase-modulated signal by comparing it with the reference signal from the stationary auxiliary antenna. The low-frequency amplitude waveform thus produced is compared with a reference waveform of the same frequency in the bearing resolver unit. The reference low-frequency wave is derived from the antenna switching circuits.

The voltage output of the bearing resolver is proportional to the sine and cosine components of the bearing angle. These voltages are applied to a bearing display amplifier, where they are used to produce a strobe waveform for the CRT indicator.

Various antenna-switching arrangements are in use. Gaseous diodes, fired by a commutating pulse, semiconductor diodes, vacuum-tube gates, or mechanical conductive or capacitive switches may be used to sequentially scan the various antenna elements.

The scanning process may sequentially scan each antenna around the circumference or may be a comparison of the phase at three or four adjacent antenna sectors; see the mechanical capacitive sampling procedure of Figure 10-53. In this scheme the sequence in sampling the antennas is as follows: 1, 2, 3, 4 ... 4, 3, ... 4, 5, 6, ...

This sequence produces an oscillatory scanning and results in a frequency deviation dependent on the direction of arrival of the signal. This deviation is greatest when the sampling sequence is at right angles to the equiphase contours, and zero when it is along a line of equal phase.

Fig. 10-53. Capacitive commutating device using "Pilgrim's Pace" scanning sequence.

10.4.1 Phase Comparison Types

During the early part of World War II a series of phase-comparison systems were developed for the Navy which had designations of CXGL (300-1000 Mc), SRD-9 (225-400 Mc), and DBO (90-400 Mc). These equipments utilized four antenna sectors, each of which covered a quadrant. A block diagram of a phase-comparison sector type is shown in Figure 10-54. The sectors could be located independently of each other in positions best suited for covering the assigned azimuth sector.

The information from the four sectors was fed sequentially, by means of a synchronously rotating capacitive commutator, to a receiver and indicator. Each of the four antenna assemblies consisted of a set of two identical antenna elements mounted in a cavity. A broadband antenna assembly for the CXGL in the frequency range 300-1000 Mc is shown in Figure 10-55. A
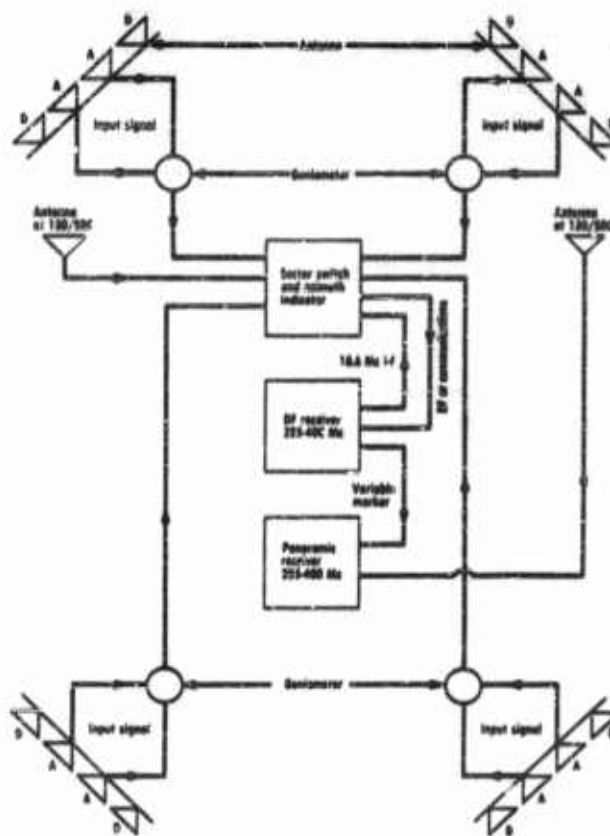
FIG. 10-54. SRD-9 phase-comparison four-sector direction finder for 225-400-Mc range.
D = dummy antenna; A = active antenna.

partition was placed between the two receiving elements, and the array was covered with a radome to form a nearly flush antenna. In back of the two antenna elements were mounted two balance boxes for converting the antenna outputs from a balanced to an unbalanced line.

The goniometer was a slotted transmission line type which inserted a differential phase shift in the individual antenna outputs, as shown in Figure 10-12. The goniometer is described in greater detail in Section 10-7.

Figure 10-56 shows a schematic of one of the goniometer and antenna-array combinations. The two antennas of the array are separated by a distance $2d$, and a signal originating at some distant point P is shown arriving at an angle θ measured from the normal to the plane of the antennas.
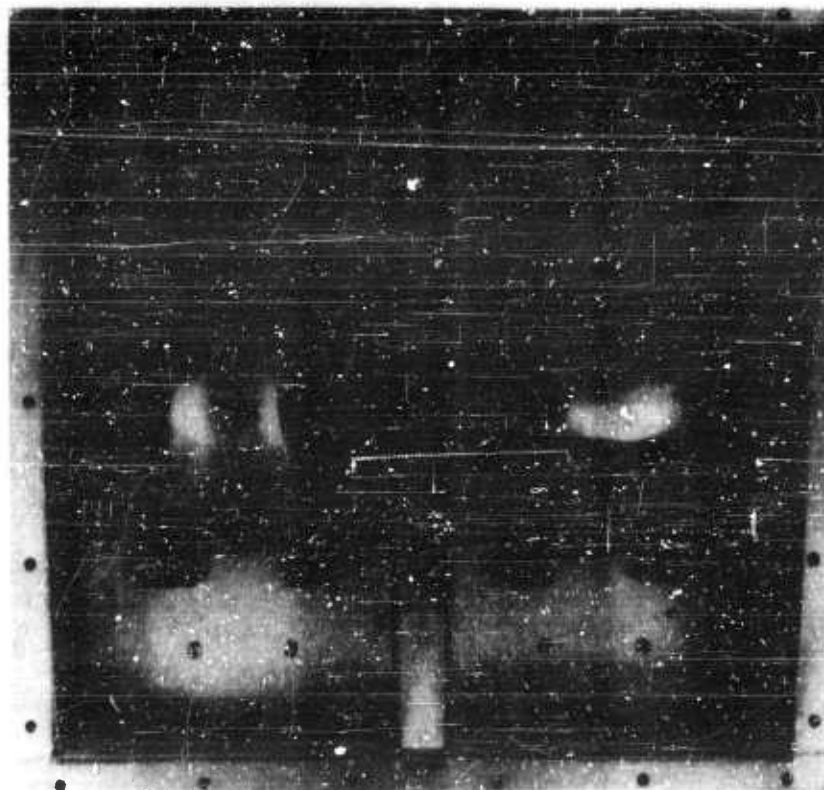
FIG. 10-55. Antenna array CXGL, 300-1000 Mc. Phase-comparison direction finder.

Since the distance to the transmitter P is very large compared to $2d$, then AP and CP are essentially parallel and antenna #1 is $2d \sin \theta$ farther from point P than is antenna #2. The field at antenna #1 is therefore retarded from that at antenna #2 by $[2d(360°)/\lambda] \sin \theta$ electrical degrees. With reference to the midpoint B, it is seen that the field E_1 at antenna #1 and the field E_2 at antenna #2 is given by

$$E_2 = Ee^{+jd'} \sin \theta$$
$$E_1 = Ee^{-jd'} \sin \theta$$

where E is a constant depending on the field strength and the antenna characteristics.

A comparison of the relative phase between V_1 and V_2 can be used to
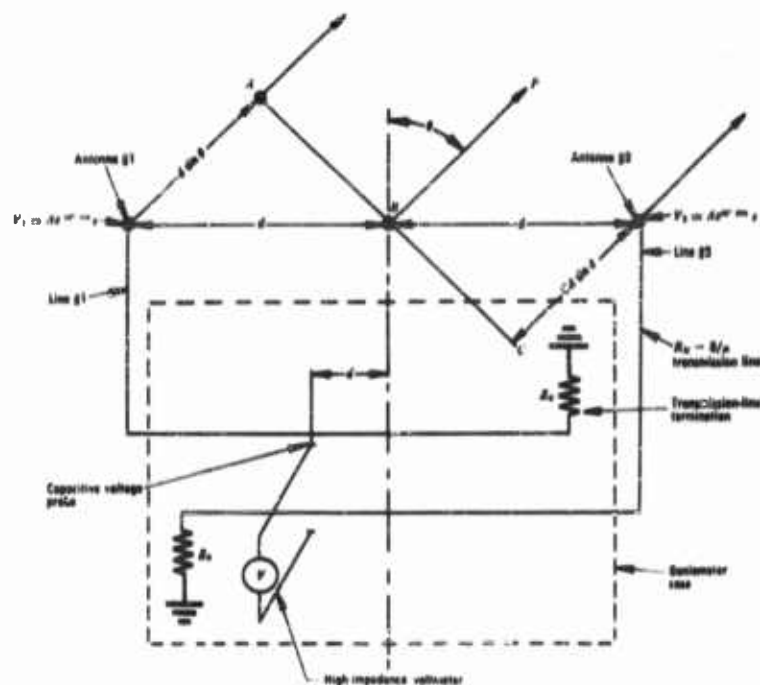
FIG. 10-55. Goniometer and antenna array. s = displacement of voltmeter from center, in electrical degrees. $2d$ = antenna spacing, in meters. $d' = d(360/\lambda)$ = antenna spacing, in electrical degrees.

determine the direction of arrival θ.

If there is mutual coupling between the two antennas, then

$$V_1 = \frac{A}{1 - M^2} (e^{-id'} \sin \theta + M e^{+id'} \sin \theta)$$

and

$$V_2 = \frac{A}{1 - M^2} (e^{+id'} \sin \theta + M e^{-id'} \sin \theta)$$

where M is a complex quantity representing the coefficient of mutual coupling between the antennas. This quantity M must be reduced to a small quantity or made independent of θ in order to prevent error from this cause.

The baffles inserted in the antenna array (Figure 10-55) were used to reduce the mutual coupling between antennas.

Figure 10-57 shows an antenna sector for the SRD-9 mounted on the U.S.S. *Leyte*. This equipment is a phase-comparison system for the 225-400-

Fig. 10-57. An antenna sector of the SRD-9 phase-comparison direction finders operating in the frequency range 225-400 Mc. The two outside antennas are dummies used to reduce the error from mutual coupling between antenna elements.

Mc range. The antenna elements are short broadband terminated V's which with their images in the ground plane form four miniature vertical rhombics. In this case, the manual coupling is not reduced but made substantially independent of θ by the use of the two outside antennas which are dummies and not connected to the system (Reference 31).

10.4.2 Interferometers

It is pointed out in Section 10.1 that all methods of direction finding utilize the phase of the field vectors. Some systems encode the phase of the field vectors by transforming the phase relations at two or more points into amplitude functions of the direction of arrival. Other methods compare the phase, by means of some instrumentation external to the antenna, of two or more points in space either simultaneously or sequentially.

Depending on the encoding process used in the collectors, some systems designated as interferometers are more properly classified as amplitude comparison methods. The Wullenweber and MUSA systems, while displaying sharp directivity patterns, are strictly phase-comparison systems.

The latter wide-aperture types were included for convenience in Section 10.3.

There are a number of direction-finding techniques which are called interferometer or diffraction-grating methods. The CXGL, a sequential phase comparison DF (Figure 10-56), may be regarded as a diffraction-grating type with each antenna being a secondary source so that the position of the first-order spectrum determines the direction of arrival. If the spacing $2d$ between secondary sources antenna #1 and antenna #2 is of the order of 135 electrical degrees or less for the highest frequency being received, then, for approximately a quadrant, only one null in the interference pattern will be encountered.

The two-antenna interferometer is by classification a sequential amplitude method and has been described in Section 10.3 in connection with reconnaissance set ASQ-18. In this method a multilobe pattern is generated by combining the outputs of two antennas separated by many wavelengths. Each antenna has enough directivity so that adjacent lobes on the multilobe pattern have noticeably different amplitudes and, as a source sweeps through the multilobe pattern, ambiguities can be resolved by identifying the lobe producing the maximum signal.

Various methods of shifting or rotating the lobes by the use of phase shifters for large base-line interferometers have been devised. These methods are directed toward separating discrete sources of signal from distributed sources, a problem often encountered in radio astronomy. This, in direction-finding practice, is equivalent to separating a signal from extraneous signals in a site where there are no large reflecting objects but rather a large number of small reflectors. Artifices such as phase switching, and rotating-lobe interferometer techniques are directed toward using more thoroughly the information concerning the field vectors available at the antennas. These methods are discussed under error-reducing techniques, in Section 10.9.

Postdetection correlation used in interferometer techniques is discussed under time-difference methods (Section 10.5) since they involve the measurement of the relative phases of components within the acceptance band. This is equivalent to determining suitable reference points on the envelope of the detected signal.

10.5 Time-Difference Scanning Methods

The phase-comparison systems described in Section 10.4 depend on the measurement of the rf phase of a signal P at antenna #1 as compared with the phase of the same signal at an antenna #2 close by (Figure 10-56). If the spacing of the antennas $2d$ exceeds $\lambda/2$, then an ambiguity will be encountered, since the same phase relation will exist for more than one angle of arrival. While larger spacings can be used, a method of resolving the ambiguities must be provided. In the Consol system of navigation (the Sonne

system used by the Germans in World War II) the ambiguity is resolved by an auxiliary low-accuracy direction finder (Reference 32).

When the structure of the signal is controllable, as in the case of navigational signals from a beacon or range, the resolution is relatively simple. For the case of an unmodulated signal whose azimuth is required from a direction-finder station, the phase of the signal repeats itself every cycle, so that, as indicated above, for large spacing of the antennas it is not known whether the values of phase being compared pertain to the same or to different cycles. In order to avoid ambiguities the measurements should be confined to the same cycle. A cycle of convenient duration may be obtained by providing a signal of frequency f_2 adjacent to f_1 and utilizing the modulation component of frequency $f_1 - f_2$. In this case the signal consists of two rf components and its phase reference is obtained at the beginning of each $(f_1 - f_2)$ cycle. Such reference is then taken at the instant at which both signals cross the time axis simultaneously. The ambiguities which exist for large antenna spacings can then be resolved since the total number of cycles between two successive simultaneous zero crossings is known and the phase-comparison measurement can be determined.

In an amplitude-modulated signal the phase of the sidebands with regard to a carrier produces an envelope change. For a pulsed signal, the relative phases of the rf components produce the amplitude changes. Therefore, a means is available for resolving the phase ambiguities inherent in a phase comparison system with wide antenna spacing. The ability to establish a reference from the envelope is determined by the bandwidth of the transmission, the signal-to-noise ratio, and the smearing of the envelope caused by propagation scintillation effects. The usual signal has a ratio of $2 \triangle B / f_o$ less than or equal to 0.1, where f_o is the spectral density point of maximum power and $\triangle B$ is the bandwidth measured from f_o to either half-power point. Thus, if the highest (f_H) and the lowest (f_L) frequency components in the transmission can be filtered out and their relative phases measured by the determination of the common zero crossing points, a phase reference can be obtained every $1/2 \triangle B$ seconds or $(150 \times 10^6)/\triangle B$ meters.

The stability of the transmitter, the characteristics of the propagation path, and the duration of the signal determine the amount of filtering that can be applied to the separation of f_H and f_L. The filtering will, in turn, determine, in the presence of noise, the probability of resolving the zero crossing to a given accuracy (References 33, 34). Since the filtering of definite portions of the transmission is generally impractical, it is customary to detect and establish the signal envelope at each of two locations and determine the corresponding difference in time of arrival from an examination of the two envelopes.

This process, which uses two receiving stations with wide separations to determine the line on which the transmitter is located, is the reverse of the loran system of navigation in which two synchronized and known transmitters and a single receiver are used to determine the locus of the receiving point. The DF technique is consequently often referred to as the inverse loran system.

If the distance between receiving points is small and about 0.01 or less of the distance to the transmitter, the geometry will be as shown in Figure 10-10.

The inverse loran methods can be classified into two categories depending on whether the system is a sequential, that is, a scanning, method, or an instantaneous, absolute method. The first of these will be discussed here; the instantaneous type will be covered in Section 10.6.

A system of direction finding based on a sequential examination of delay to determine the difference in time of arrival of a signal at two spaced antenna arrays is shown in Figure 10-58.

Two antenna arrays are separated by an approximate distance

$$d = v/\triangle B$$

where $\triangle B$ is postdetection filter bandwidth and v is the velocity of light.

Each antenna is connected to a receiver tuned to the desired signal. After detection and filtering, the output of one filter F_1 is delayed by an adjustable amount and fed into a correlator. The output of each filter is fed into the other input of the correlator. The output of each filter is also rectified and filtered to obtain the rms values V_1 and V_2, respectively. These rms values are multiplied together and the result is utilized to normalize the correlator output. The quantity

$$K = \phi_{12}/V_1 V_2$$

is called the normalized correlation coefficient.
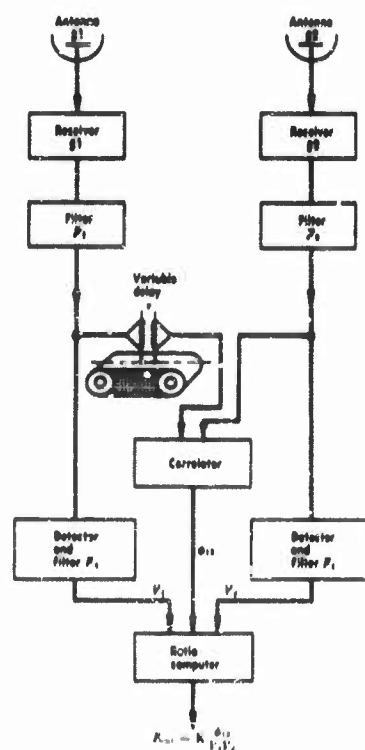
Fig. 10-58. Postdetector correlator direction finder.

The delay τ is now adjusted until K is a maximum. This delay τ_{max} is the latency in time of arrival of the signal at one antenna with respect to the other.

The correlator is described more fully under instrumentation in Section 10.7.

A second DF technique used on wideband signals such as radar pulses is shown in Figure 10-59. Antennas #1 and #2 are each connected to a crystal detector followed by a video amplifier. The output of each crystal video receiver is connected to a delay line. The vertical plates of a cathode-ray oscilloscope are connected to a switch which selects various taps on a delay line, and the horizontal plates of the oscilloscope are connected to select the taps on the other delay line. These connections are sequential so as to increase the delay for the signal from one antenna as the delay for the signal from the other is decreased. The taps selected yield the closest in-phase configuration for the Lissajous figure on the oscilloscope face.

Figure 10-60 shows some figures obtained with this technique.

FIG. 10-59. Crystal-video inverse loran direction-finding technique.

10.6 Instantaneous Types

In order to obtain a bearing on a signal which is of short duration it is desirable to use DF techniques which will yield a bearing in the same time interval required for the detection of the signal.

Direction finders which accomplish this are called instantaneous types. Rapid-scanning techniques are not included in this category since such techniques result in a wider bandwidth requirement and therefore in a deterioration of signal-to-noise ratio.

10.6.1 Instantaneous Comparison of Amplitude

The instantaneous comparison of the amplitude of a signal from the two pairs of antennas forming an Adcock array may be used to determine the direction of arrival. The output from each Adcock pair is fed to separate re-

Fig. 10-60. Lissajous figures for delay line, inverse loran system; 100-foot base line against X-band radar.

ceivers. The output of one receiver is applied to one set of deflection plates of a cathode-ray-tube indicator and the output of the other receiver to the other set of deflection plates. The slope of the in-phase Lissajous figure indicates the direction of arrival of the signal.

Quadrant identification must be obtained from a measurement of the relative rf phases of the antenna outputs or from a separate "sense" antenna. Often a third receiver is used for this purpose, the output of which is combined with the output of the two main-channel receivers at i-f.

Since in systems of this type the amplitude comparison is on an instant-to-instant basis, any distortions of waveform due to phase shift in the rf or video circuits will produce an error. An asymmetrical amplification of sidebands will also result in a bearing error.

The design of balanced receivers provides for stage-by-stage equalization with ganged controls and paralleled trimming auxiliaries. Frequent calibration is necessary since it is difficult to maintain equal receiver performance over cyclical temperature and humidity changes. Vibration being considered as an unreasonable complication, equipments with such balanced receivers are not designed for mobile use.

A DF system in which there is a comparison of the signals received simultaneously by two or more antenna arrays having substantially orthogonal directive patterns may be classified as an instantaneous amplitude-comparison type.

Such a system with carefully balanced receivers was used by Watson-Watt and Herd (Reference 35) for obtaining instantaneous bearings on electromagnetic radio sources of atmospheric origin.

A system using a four-element Adcock array is shown in Figure 10-61.
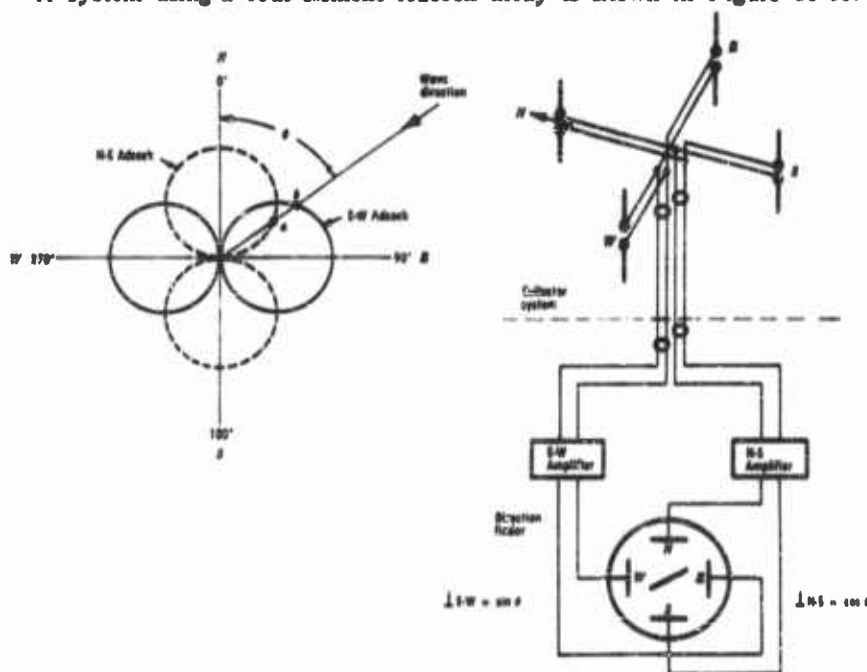
Fig. 10-61. Direction-finding system using a four-element Adcock array. In the antenna polar pattern, $\theta =$ the angle between the azimuth of direction of arrival and a line perpendicular to the line joining the two antennas.

Each of the opposite pairs of the elevated H array is connected to a receiver, the N-S pair being connected to receiver #1 and the E-W pair to receiver #2. The receivers must have a gain for weak signals of about 160 db, and must maintain this equality of gain for various signal levels and at all frequencies within the tuning range.

An instantaneous type of DF equipment using crossed loops is shown in Figure 10-62. This equipment, designated as the AN/TRD-7, covers the 60-300-Mc range and was intended primarily for the location of jammers

Fig. 10-62. Direction-finding using crossed loops.

operating against radio-fuzed weapons. The frequency range is covered in two bands, 60-150 Mc, and 150-300 Mc. A central sense antenna provides a signal which, when amplified and combined at i-f with either of the two similarly amplified outputs of the loops, produces a cardioid pattern. The quadrant in which the signal originated can thus be identified.

The military did not use many of these equipments because of the unusually high skill required by field personnel to keep them in operating condition. Typical field tests showed that elaborate aligning procedures had to be repeated at intervals of less than one hour in order to obtain useful performance.

An alternative method of amplifying the signals from the orthogonal-pattern antenna arrays is to identify each signal by a characteristic modulation and then employ a common receiver for all channels. This method is commonly called the single-receiver system.

The single-receiver system is particularly applicable to the usual amplitude-modulated communication-type signals. Although the single-receiver system will, in general, be more susceptible to error when more than one signal is being received at one time, its stability and reliability more than make up for this deficiency.

Deliberate jamming, however, may cause considerable difficulty with the single-receiver system, and the instantaneous feature of the system may not be worth the loss of a certain amount of immunity to jamming.

The method of identifying the signal which is fed from each of the two Adcock pairs may be accomplished in a variety of ways.

In one system (Reference 36) each of the signals from the Adcock pairs is modulated by a separate audio-frequency while the center or omnidirectional antenna is modulated by a third audio-frequency. After amplification and detection, the three audio components are separated by three filters. The audio-frequencies which were used to modulate the directional elements are translated to the same frequency as the third or omnidirectional modulating signal.

The two directional components may be used, by arranging them in time quadrature, in generating a propeller pattern on an oscilloscope screen. The component from the omnidirectional antenna may be added, with a phase adjustment, to the quadrature voltages to produce an asymmetry and resolve the 180° ambiguity.

A version of the single receiver system using two audio modulating frequencies has been used to improve the indication by making the rectified carrier level substantially independent of azimuth. In this modified system (Reference 37) the output of each directional pair is modulated by a different audio-frequency by means of a balanced modulator. The carrier is suppressed by the modulators and restored after amplification by the addition of the signal, properly phased, from the omnidirectional center element.

The output of the receiver consists of two audio-frequency components whose relative amplitude is a function of the bearing of the signal.

A further modification of the two-tone system led to the use of sine and cosine components of a single audio tone. The phase of the audio tone, after amplification, demodulation, and filtering, is then a measure of the direction of arrival of the signal. The bearing is obtained by comparing the phase of the filter output with the modulating audio-oscillator, after suitable calibration. The system is similar to the instrumentation proposed and used in the quadrature-field goniometer (References 38, 39).

Although the use of the orthogonal functions of a single frequency is more economical of channel capacity, nevertheless, if phase shifts are encountered in the receiver, producing a phase change in the modulation frequency, an error equal to that phase change is encountered for all directions of arrival. The use of two frequencies, on the other hand, produces an error of quadrantal form which is zero for the four quadrature directions and of the form

$$\theta_{an} = (\cos 2\theta)(1 - \cos \alpha)$$

where α is the modulation envelope phase shift, and θ is the true bearing of the signal.

A single receiver, selective modulation, instantaneous Adcock DF system is shown in block diagram form in Figure 10-63. The output of
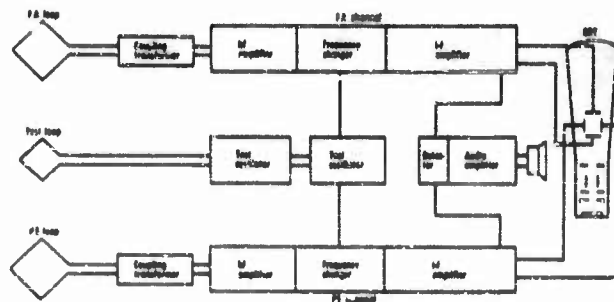
Fig. 10-63. Instantaneous Adcock DF system using a single receiver with selective modulation.

each of the two diametrically opposed pairs E-W and N-S is connected to the E-W and N-S balanced modulators, respectively. The output of the center or omnidirectional antenna is amplified and combined to furnish an azimuth independent carrier in place of the suppressed carriers in the output of the modulators. The combined signal is then amplified, demodulated, and filtered by the receiver. The output of the receiver is amplified by an audio amplifier and connected to the signal coil of the DF indicator.

A pair of fixed space-quadrature field coils is provided in the indicator, the N-S pair being excited by a current i_1 from oscillator f_1 and the E-W by a current i_2 from the oscillator f_2.

The in-phase components of each frequency exert torques in space quadrature so that the pointer assumes a position at which the average value of these torques is zero.

With a system of this kind the instrumental accuracy (neglecting siting and polarization errors) can be such as to produce errors of less than $\pm 4°$ at the least favorable frequency and azimuth positions. Of this, approximately half of the error is of quadrantal nature and with some labor at proper calibration can be reduced substantially.

10.6.2 "Wide-Open" Systems

For fixing the location of radar signal sources, it is often necessary to employ special techniques. The signals under these conditions are pulsed, short-duration bursts, modulated from pulse to pulse by the antenna rotation, and may have the additional feature of pulse-to-pulse frequency shift.

The range of a radar set is given by the simple relation:

$$R_m = 0.1146(KP_P)^{1/4}$$

where K is a constant depending on a large number of factors including pulse

width, wavelength, effective target area, losses, and antenna gain, while P_P is the peak transmitter power.

Since the range within which the detection and DF operations can be performed is a function of the square root of P_P, less sensitive receivers may be used for these operations than for the radar function.

A wideband crystal video receiver is generally adequate for line-of-sight airborne intercepts aga ·s. ground radars at altitudes up to about 40,000 feet.

For longer ranges, encountered at higher altitudes or when interception is necessary on weak minor lobes, an rf amplifier, usually using a TWT, is interposed ahead of the crystal for increased sensitivity.

A crystal video amplifier is shown in Figure 10-64.

FIG. 10-64. AN/ASQ-23 crystal video amplifier and count-down unit.

Since crystal video receivers can be made small and compact, it is possible to design instantaneous types of direction finders using a number of such receivers each of which is associated with a horn or other directive antenna.

The antenna need not be centrally located but can be placed at convenient positions on the periphery of the vehicle or platform so as to view a relatively unobstructed sector.

An instantaneous amplitude-comparison system of this type may have a completely separate channel for each antenna sector or, since the duty cycles of pulsed radar transmissions are low, after preliminary amplification, the outputs of each horn may be sequentially amplified by introducing a different delay for each output.
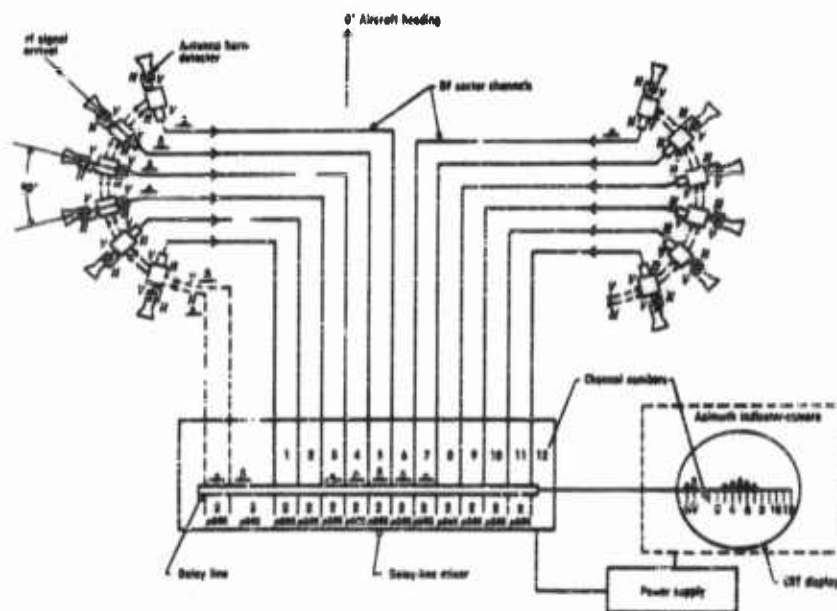
Fig. 10-65. Simplified block diagram for one-third of an AN/APD-4 system.
(CONFIDENTIAL)

Figure 10-65 shows a simplified block diagram of an operational instantaneous direction finder, the AN/APD-4, operating in the frequency range 1000 to 33,000 Mc. This frequency range is covered in three bands.

The equipment uses twelve horn antennas for each of the three frequency bands. Each horn antenna has associated with it two probes, one for vertical and the other for horizontal polarization. Each probe is connected to a preamplifier to bring the level above the noise level of the subsequent amplifiers by an amount equal to or greater than the loss in the delay lines which follow it.

In order to avoid having to provide a complete amplifier including a high-level output stage with each horn output, a delay line is provided which introduces an incrementally greater delay in each successive antenna output. The result is to produce, from a single pulse signal, a series of pulses sequentially staggered in time which can then be fed through a single amplifier.

The signal is displayed on a cathode-ray tube and recorded photographically, as shown in Figure 10-66.

The display is arranged so that a single pulse produces, as a first signal, a superimposed output of the horizontally polarized component outputs of

Sawtooth pr rate
(low)

Sawtooth pr rate
(high)

Moving-spot pr rate
(low)

Signal polarization

Gyrosyn

Clock

X-band signal
vertically polarized
(low rotational rate)

2-sec markers

S-band signal
horizontally polarized
0.5-sec pulse width
(high rotational rate)

Signal headings

Moving-spot pr rate
(high)
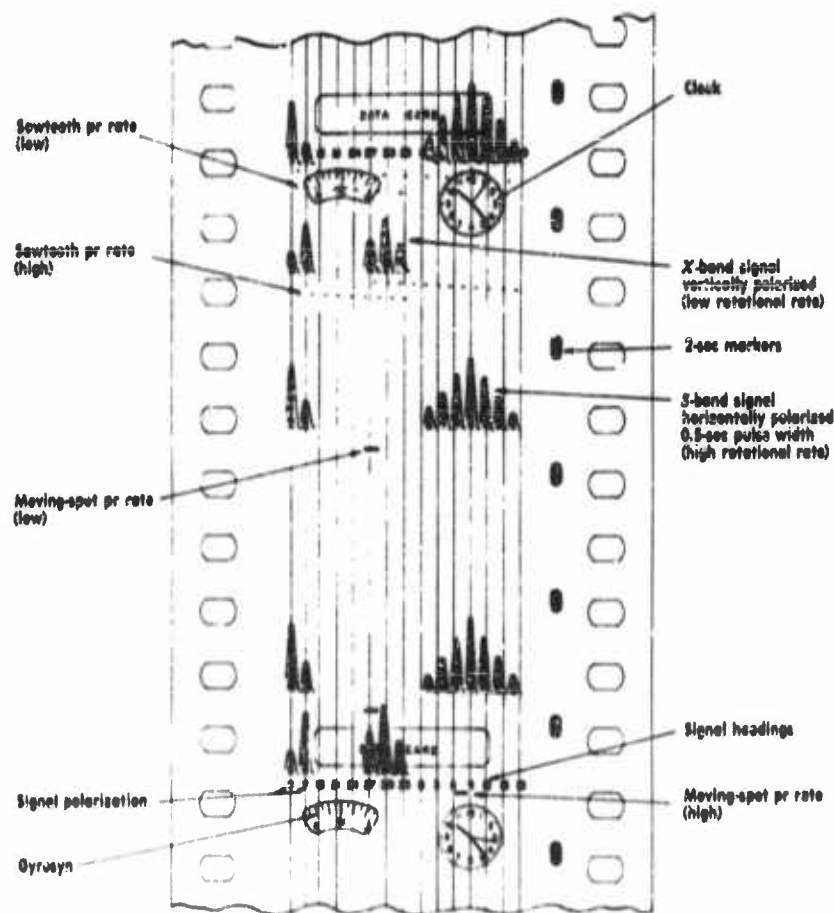
Fig. 10-66. Recorded display of AN/APD-4 wide-open intercept system. Clock, signal heading, data card, and gyrosyn flashed on every 10 seconds. (CONFIDENTIAL)

all the antennas. T. second signal, moving from left to right on a horizontal sweep, is the summed output of the vertically polarized signal components. The succeeding signal, in terms of increasing delays, represents the output, to the same pulse, of the various horns.

The direction of arrival of the signal is given by the position of the maximum of the envelope representing the various antenna outputs.

Since the directive pattern of each antenna sharpens with increasing frequency, the width of the envelope is a measure of the rf of the signal.

A second gun in the cathode-ray tube is intensity modulated by the signal and produces a series of dots as shown in the display, indicating the pulse-repetition rate.

From the display, therefore, the direction of arrival, pulse shape, polarization, approximate rf, pulse-repetition rate, antenna pattern, and antenna rotational rate may be observed.

Figure 10-67 shows such an instantaneous DF equipment installed in an

FIG. 10-67. Wide-open intercepting system installed i nan RB-50 aircraft.

RB-50 aircraft. The antennas are located at the wing tips so as to be as free as possible from errors due to reflection from the aircraft surfaces.

The various wide-open intercept and DF techniques operating in the microwave region use, in the main, either horns or Luneberg lens configurations for azimuth resolution. The essential difference is in the method of signal processing. The AN/APD-4, as indicated above, uses a wideband photographic recording with the signal processing relegated for subsequent ground analysis.

The AN/DLD-1 equipment is a wide-open system also using crystal video receivers, and covering the frequency range 1-40 kMc. Luneberg lenses similar to that shown in Figure 10-49 are used in this equipment for the 1-9.1 kMc frequency band and an array of fifteen horns for the 9.1-40 kMc band. The AN/DLD-1 antennas or probes in the Luneberg lens channelize the azimuthal plane into fifteen sectors. The output of each sector antenna or probe is detected by a crystal detector. The resultant fifteen video outputs are fed into a simple analyzer which compares the outputs of adjacent channels. If a pulse is predominantly received in a single sector, the direction of arrival is specified as the direction of the sector orientation. The direction of arrival is recorded under these conditions as being within $\pm 6°$ of the center of the sector. If pulses are received on two adjacent sectors and their amplitudes are judged equal within some design limit such as ± 3 db, then

the direction of arrival is recorded as that corresponding to the angle which lies between the adjacent sectors and within $\pm 6°$ of this value.

Thus, by arithmetic operations on the relative intensity of the signals detected in the original fifteen directional antennas, thirty directions of arrival are deduced and each incoming signal is recorded as having one of these thirty directions.

10.6.3 Instantaneous Comparison of Phase

The line of direction of a transmitter may be found without scanning by comparing, on an instantaneous basis, the phase at two or more points, provided that the sampling points are sufficiently close to avoid ambiguities. Figure 10-68 shows a diagram of a "sum-and-difference" DF. If the rf is
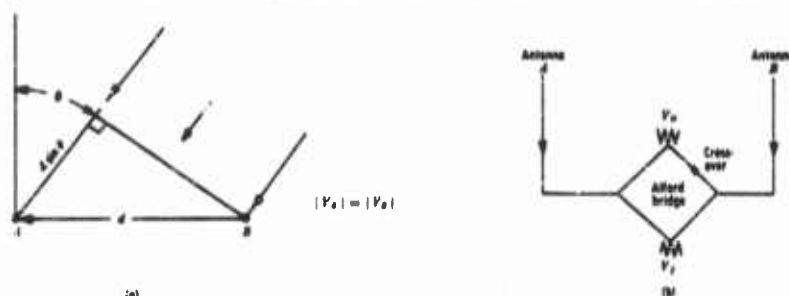
Fig. 10-68. Instantaneous comparison of phase.

not known accurately but sufficiently well to establish that the sampling points are closer than $\lambda/2$ apart, then the absolute value of the sum of the field at points A and B is:

$$\left| V_s \right| = V \left| 1 + \epsilon^{-j\,2\pi\,(d/\lambda)\,\sin\theta} \right| = 2V \left| \cos\left(\frac{\pi d}{\lambda} \sin\theta \right) \right|$$

This sum is available at the #1 terminal of an Alford bridge. The difference is available at terminal #2 by means of the crossover or phase reversal marked by an X in an arm of the bridge. The absolute value of this difference is:

$$\left| V_D \right| = V \left| 1 + \epsilon^{-j\,2\pi(d/\lambda)\,\sin\theta} \right| = 2 \left| \sin\left(\frac{\pi d}{\lambda} \sin\theta \right) \right|$$

Taking the ratio of the sum and difference:

$$\left| \frac{V_D}{V_s} \right| = \tan\left(\frac{\pi d}{\lambda} \sin\theta \right)$$

or the angle of arrival is

$$\theta = \sin^{-1}\left(\frac{\lambda}{\pi d}\tan^{-1}\frac{V_D}{V_s}\right)$$

The direction of arrival can of course be obtained by any number of sampling points judiciously selected so as to avoid ambiguities and also the necessity for knowing *a priori* the frequency of the signal. In the ideal case, some of this information would be redundant for determining the direction of arrival if the frequency is known. Many interesting schemes can be devised, particularly when phase-perturbing influences are present and their effect is to be minimized.

10.7 Measurement Techniques—Goniometers

The antennas associated with a direction finder may be considered, as previously indicated, to be exploration or sampling points in space. The output signals from these antennas are connected to a device which combines them so as to yield a signal parameter which varies with the direction of the signal source. If the device combines or encodes the outputs sequentially and with different arrangements of phase, amplitude, or delay so as to produce an output which is characteristic and dependent on the direction of arrival, it is called a radio goniometer.

A goniometer for use with crossed loops in the Bellini-Tosi arrangement is shown in Figure 10-69.
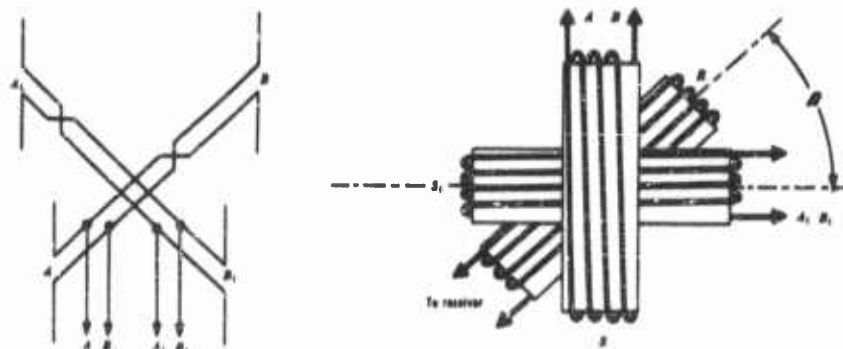
Fig. 10-69.   Principle of the inductive goniometer.

It is assumed that the geometric center of the antenna array is the phase reference for the system; that is, the phase of the rf at the output of each loop is the same as would have existed at the common vertical axis for the loops.

A further assumption is made that the movable coil of the goniometer does not disturb the distribution of flux from either stationary coil for any position.

To rigorously meet these conditions over a wide band of frequencies and without elaborate compensation, the antenna would have to extract negligible power from the field; similarly, the goniometer moving coil could not load or absorb power from the stationary coils.

With these and other simplifying conditions, each loop, or in the case of an Adcock system, each pair of antennas, produces a figure-eight pattern, the two figure-eight patterns being in space quadrature.

The currents I_{AB} and I_{A1B1} in the stationary coils of the goniometer are assumed to be proportional to those of the corresponding antennas A, B and A_1B_1. Hence:

$$I_{AB} = I_{max} \cos \theta$$

$$I_{A1B1} = I_{max} \sin \theta$$

where θ is the angle of arrival of the signal with respect to the plane of the AB loop or antenna pair.

The field coils are arranged in space quadrature with each other; therefore, the current I_R induced in the rotating goniometer coil is, in view of the simplifying assumptions made above:

$$I_R \alpha (I_{AB} + jI_{A1B1}) \alpha I_{max} (\cos \theta \cos \beta + j \sin \theta \sin \beta)$$

where β is the angle of the rotating coil with respect to the plane of the AB loop or antenna pair. The current has a maximum value I_{max} when $\theta = \beta$.

The movable coil of the goniometer is therefore rotated for either a maximum or minimum output in order to determine the direction of arrival of the signal, the minimum occurring 90° off the maximum, that is, where $\cos \theta = \sin \beta$.

Figure 10-70 is a view of an inductive goniometer for the frequency range 2 to 8 Mc. The coil in the center of the shaft is the output device for coupling to the receiver input.

The goniometer may be designed to utilize capacitance for the coupling between the fixed branches and the movable pickup element. Such a goniometer may consist of two pairs of fixed plates at right angles to each other, as shown in Figure 10-71. The two movable plates C_R—C_R are connected to the input of a receiver.

Many precautions must be taken to avoid fringing and variations in the pickup output as the device is rotated. The capacitive coupling is often reduced to a minimum or swamped by larger fixed shunt capacitance in order
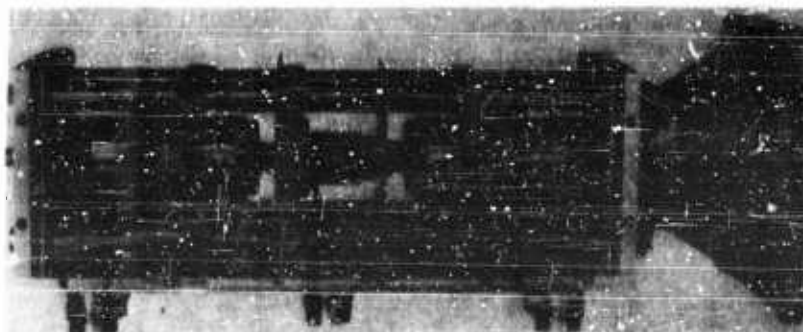
Fig. 10-70. Inductive goniometer for frequency band 2-8 Mc.
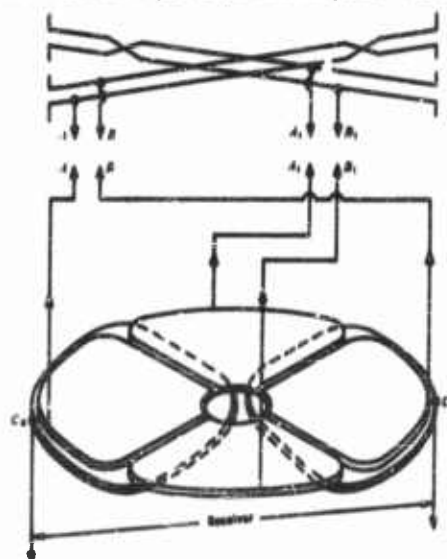
Fig. 10-71. Capacitive goniometer used with Adcock collectors.

to prevent modulation of the signal in other than a fundamental sinusoidal manner.

Figure 10-72 shows a capacitive goniometer for operation in the frequency band 100 to 160 Mc. Capacitive coupling is also used to couple from the movable rotor plates to the fixed output terminals which connect to the receiver. The capacitance is obtained by concentric cylinders, properly shaped. Often the goniometer is intentionally designed to employ both inductive and capacitive coupling. In this manner the operation is made reasonably efficient over a wider frequency band.

Goniometers using transmission lines can be designed for use with sequential-phase-comparison types such as that shown in Figure 10-56.

Fig. 10-72. Capacitive goniometer operating from 100 to 160 Mc.

Fig. 10-73. Transmission-line goniometer for sequential-phase direction finder (300-1000 Mc).

A goniometer of this type is shown in Figure 10-73 for the frequency range 300 to 1000 Mc. The design of such a goniometer is an interesting problem. Transition sections must be provided at positions where the geometrical configurations change, so as to avoid fringing and distortions of the field and thus introduce discontinuities and reflections.

The goniometer consists of two circular transmission lines slotted to permit tight coupling to a pair of capacitive probes. Each transmission line is fed by an antenna and terminated by its characteristic impedance. The transmission lines are arranged so as to be fed at opposite ends, so that a movement of the probe in one direction will introduce a differential phase shift between the two line inputs, that is, a positive phase shift in one line, and a negative phase shift in the other.

The capacitive probes are connected to a quarter-wave balun which measures the difference in phase at each probe point. The balun outside conductor is made as small as possible so that its impedance with respect to the

surrounding cavity is as high as possible in order to minimize loading at the null point over as wide a frequency range as possible. For the balun shown in Figure 10-73, silver surgical tubing was used as the outside conductor and an insulated wire a few mils in diameter inserted as an inner conductor.

The power extracted by the probe becomes a minimum when the equiphase positions are found, so that the device, when operated as a null indicating goniometer, introduces an error of less than 1°.

Goniometers have been made for the lower frequency ranges (225 to 400 Mc) having errors, at the null position, of less than 1°. Such a goniometer, rotatable at 1800 rpm, is shown in Figure 10-74. Four goniometers of this type are used with the phase-comparison DF previously shown in the block diagram of Figure 10-54 and in the photograph of Figure 10-57.

Goniometers using gaseous plasma have been used at the microwave region. Figure 10-75 shows a phase-comparison goniometer for use at X-band. The

Fig. 10-74. Ionized-gas goniometer for sequential-phase direction finder (225-400 Mc).

Fig. 10-75. Ionized-gas goniometer for sequential-phase direction finder (variable d-c applied through terminals A and B).

phase velocity of the signal in the waveguides is varied by varying the ionizing current.

An ionized gas contains free electrons, and the presence of these electrons

changes the phase velocity of an electromagnetic wave passing through the medium. See Section 10.2.

The phase velocity of the wave in the guide is increased by a factor dependent on the number of electrons per unit volume and as an inverse function of the frequency of the signal.

If two waveguides, each connected to a horn, are in turn brought together with a 180° differential electrical length in one guide, then a null is obtained at a probe located at the junction.

When an ionized gaseous atmosphere is provided in one waveguide of the configuration, the phase velocity is increased in that arm. The null of the combined antenna pattern is thus shifted in the direction of the guide containing the ionized medium.
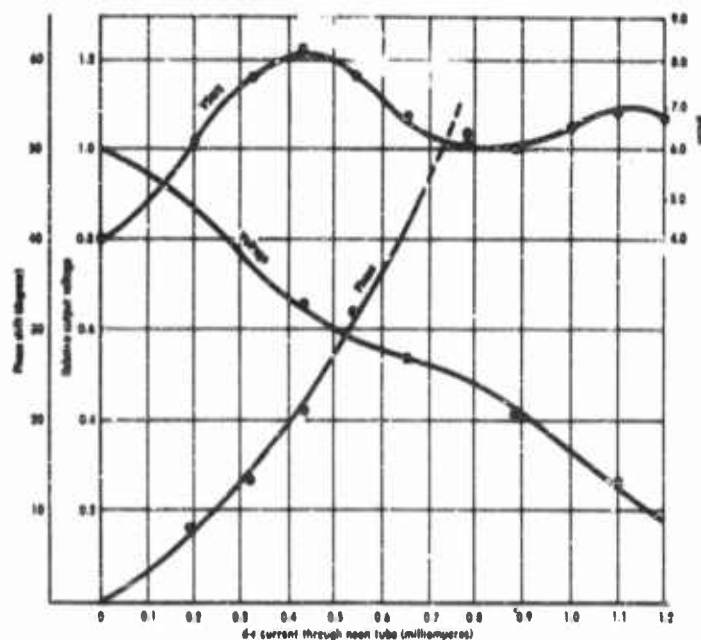
Fig 10-76. Characteristics of neon-loaded waveguide (RG-52/V), 18 inches long. Frequency = 9540 Mc.

Figure 10-76 shows the characteristics of a neon-loaded RG-52/V waveguide. By varying the ionizing current, a phase shift can be produced which causes the null of the combined antenna pattern to shift. A period movement of the pattern can be obtained by a periodic variation of the ionizing current, and thus a sector-scanning phase-comparison direction finder can be assembled.

To obtain high scanning speeds, electronegative gases are introduced into the neon tube contained within the guide. These gases cause a more rapid recombination of the ions and electrons and thus hasten the deionizing process.

At radio frequencies where traveling-wave tubes are practical, a suitable phase-comparison goniometer may be made by substituting a TWT amplifier for each waveguide, as shown in Figure 10-77.
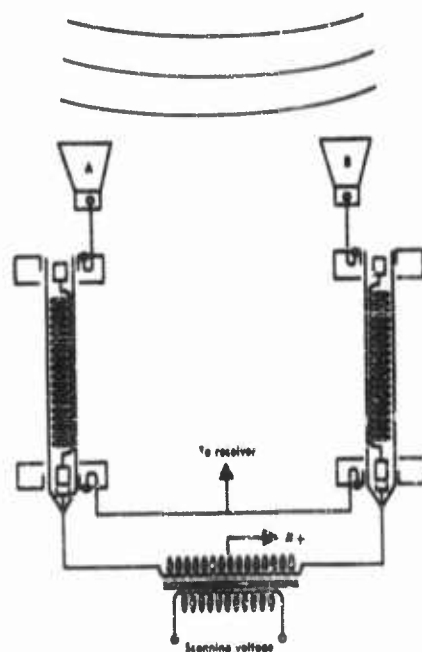
FIG. 10-77. Traveling-wave-tube goniometer.

The phase shift in each traveling-wave tube is, within limits, substantially linear with beam current. By applying an alternating voltage in push-pull fashion to the collector elements, a differential phase shift can be introduced in the rf outputs from the horns A and B. The result is that the null in the antenna pattern is caused to move in azimuth. The output of the goniometer is connected to a receiver and then applied to a suitable indicator.

At 3000 Mc, differential phase shifts of over 1000° can be obtained with such arrangements. Since under such conditions there is a variation of gain of from 5 to 10 db per arm, it is better to operate on the null rather than a maximum of the resultant antenna pattern.

10.8 Display Techniques

To determine the direction of arrival of a signal, a means must be provided for the observer to determine how the signal is being modified by the direction characteristics of the antenna array. Thus, in direction finders using a sequential amplitude-comparison method such as the rotation of a loop, the earliest methods used to provide an aural indication to the observer were by means of a headset. In the event the signal had no audible modulation, a local oscillator was provided and so adjusted as to produce an audible beat frequency with the carrier of the incoming signal. Aural methods are useful when the signal is of sufficiently long duration to permit a careful determination of the null and when the signal-to-noise ratio is high. If a local oscillator is used, the audible frequency is best adjusted to fall between 1000 and

6000 cycles. In this frequency region, the minimum audible field averages about 55 db below 1 dyne per square centimeter of sound pressure. The ear is able to selectively perceive a steady tone in a high ambient noise background since noise disturbances falling outside of a frequency area close to the steady frequency do not interfere with the perception of the tone. Indeed, it is difficult to develop an electronic instrumentation which approaches the sensitivity and has the signal-to-noise enhancement processes found in the human auditory system.

l systems are also used with the beam-switching types of direction finders. One such application is the direction finder in which a cardioid is flipped through 180° so that its minimum becomes a maximum. This switching can be done at such a rate that the pattern is in one position during the "on" times of the character A in the Morse code and on the other side during the "on" times of the Morse character N. When the switch-lobe direction finder is manually rotated so that the axis about which the lobes are switched is pointing toward the transmitter, the intensities of the A and the N characters become equal and the resultant interlock produces a steady tone. This system is particularly useful on a steady signal when the noise background is high and the lobes being switched have their maximum sharpness (rate of change of pattern with azimuth) along the interlock direction. This is generally the case when endfire types of antenna arrays are being used

Many other interesting methods of indication can be devised using aural indication. However, care must be exercised to utilize to best advantage the physio-psychological attributes of the human auditory system (Reference 40).

The use of a meter which reads the signal level at the output of a DF receiver can be used to determine a null position of the receiving antenna y. By smoothing out the input to the meter, rapid fluctuations due to extraneous noise are filtered out. In the manually operated switched-lobe direction finder, the meter may be made to read plus or minus from a zero position and thus indicate in which direction the antenna array must be rotated to approach the null position.

In the automatic direction finder, as introduced by Busignies (Reference 39), a goniometer is generally rotated at a constant speed by a synchronous motor. The goniometer shaft also drives a permanent magnet alternator for the usual closely spaced Adcock antenna arrays. The output of the goniometer is a sinusoidally modulated signal. The modulated sine wave has a frequency equal to the rate at which the goniometer is being rotated and is therefore at the same frequency as the signal generated by the permanent magnet alternator. The relative phase of the alternator signal to that of the goniometer-modulated signal indicates the direction of arrival. Normally, the goniometer is rotated at 1800 rpm; the receiver output will therefore be

modulated at a 30-cycle rate. This 30-cycle wave is filtered and applied to a phase-comparison device where the phase of this filtered 30-cycle wave is compared with the 30-cycle wave generated by the permanent magnet alternator rigidly coupled and rotated synchronously with the goniometer shaft.

The phase-comparison device may be one of many types. If the alternator output signal is passed through a phase-shifting network, then the phase-shifted 30-cycle frequency and the receiver-filtered output can be applied to the two inputs of a wattmeter type indicator. The phase shifter is then varied until the wattmeter output is a maximum. The reading of the phase shifter for this condition indicates the direction of arrival of the signal. This system is shown schematically in Figure 10-78.
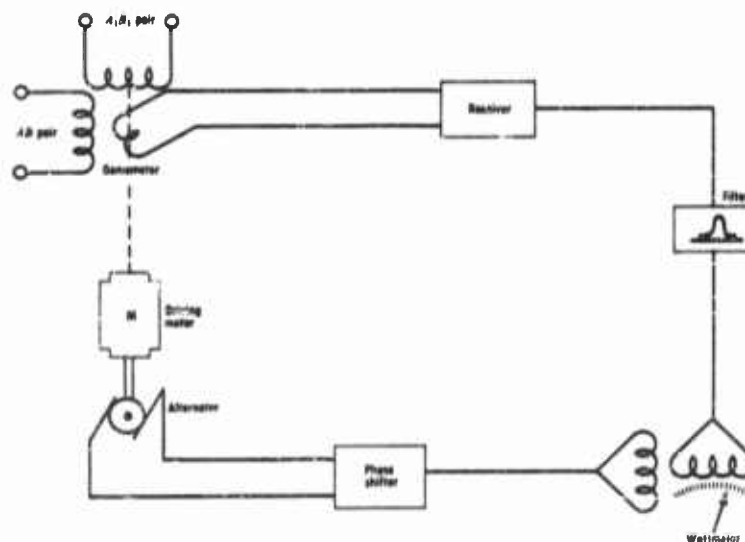
Fig. 10-78. Wattmeter used with goniometer as indicator.

Other phase-comparison methods may be readily devised. A stroboscopic method is used in the Type 14 direction finder, as shown in Figure 10-79. In this direction finder, the receiver output, after having been filtered, is clipped and differentiated so as to mark the zero crossing of the wave. The peaked signal which results is amplified and applied to a spiral neon lamp located behind a frosted glass dial face. Between the spiral neon lamp and the frosted glass is a disk with a radial slot, driven by a motor synchronized with the goniometer rotation. In operation, a radial illumination is seen on the dial face, indicating the direction of arrival of the signal.

The cathode-ray tube has a number of advantages when used as an indi-
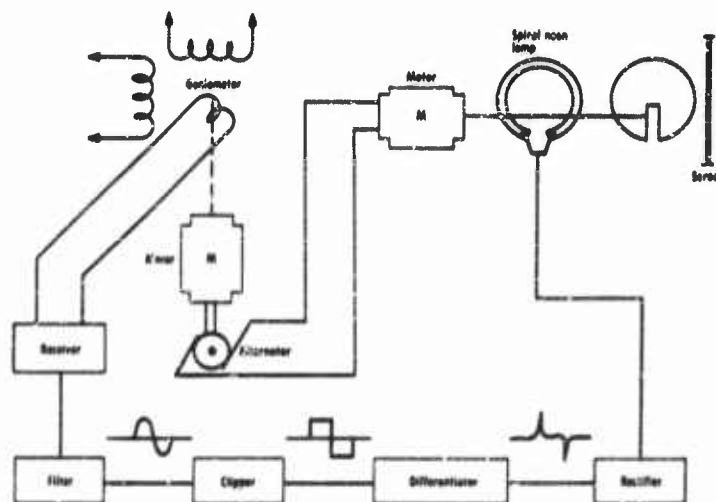
Fig. 10-79. Neon flash indicator used with goniometer.

cator. (1) It is a wideband device; (2) it is a storage medium with a short
time memory; (3) it permits the mapping of a signal so as to utilize the
physiopsychological optical attributes of the observer; that is, the display
may be tailored to provide a favored signal for pattern-recognition purposes.
The cathode-ray tube has been used as an indicator for many years, but few
of the systems have utilized all of the three factors noted above. In one of the
earliest applications of the cathode-ray tube, it was used as an indicator
device with the Watson-Watt instantaneous amplitude-comparison method
already described. The cathode-ray indicator was subsequently used by other
equipment manufacturers to compare the output of space-quadrature antenna
patterns. In one form, a loop was rotated at 1800 rpm. The motor which
rotated the loop also drove an alternator synchronously with the loop rota-
tion. The alternator in this case was a two-phase type producing a sine and
a cosine voltage. These two voltages were applied to the vertical and hori-
zontal plates of a cathode-ray tube, respectively, so as to produce a circular
pattern. The amplitude of the sine and cosine voltages was modulated by
the output of the receiver connected to the rotating antenna. In this manner,
the spot moved in and outward from the center in conformity with the re-
ceiver output, producing a propeller pattern, as shown in Figure 10-80. The
operator could, by using such a display, select the bearings associated with
the strong signal portion of a fluctuating bearing and thus obtain more ac-
curate indications.

Many other types of indicators have been devised, both mechanical and

Fig. 10-80. Visual indicator showing typical propeller pattern.

of the cathode-ray type. However, a few general considerations are important whether a human observer obtains the bearing from a cathode-ray tube, a meter, an aural system, or the output of a computer which has operated on the digitized instantaneous values of the signal, amplitude, or phase. The bandwidth of the receiver before final detection should be as narrow as possible without rejection of the signal and within the frequency stability not only of the transmitter but also of the receiver. The postdetection bandwidth should be sufficiently wide to permit fluctuations in the bearing of the signal to be recorded or observed. Since most bearings over the period of observation do not have a zero mean value, the best bearing can only be obtained by observing the manner in which the direction of arrival varies from instant to instant. An indicator, therefore, which provides the observer with a recording or memory of the bearing over a period of time permits the best interpretation. Much of the success in the operation of a direction finder depends upon the skill and intuition of the operator.

10.9 Position Fixing

To determine the position of a transmitter requires that at least two bearings be obtained, each from a different azimuthal position relative to the signal source. If the transmitter is moving with respect to the DF stations, it is best that these bearings be obtained simultaneously.

The accuracy with which the position of the transmitter can be determined is improved by utilizing more than two DF stations, as shown in Figure 10-81. These bearings generally do not intersect at the same point and, therefore, some method must be utilized to determine the most probable intersection. Errors which introduce inaccuracies in each line of direction determination may be due to several sources.

A major source of error is caused by propagation anomalies when dealing with hf signals in which the propagation is by means of the sky wave. Such errors arise from the fact that the ionosphere is not perfectly uniform in structure. As a result, lateral deviations of the rays take place. In addition to this, as pointed out previously, there are polarization changes as the wave is reflected from the ionosphere.

At higher frequencies, above 60 Mc, scintillation of the apparent direction of arrival of the wave may occur due to contributions to the field at the re-
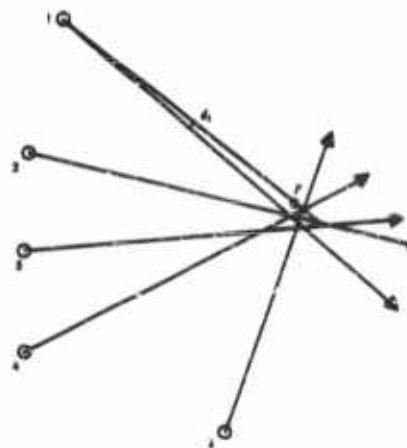
Fig. 10-21. Direction-finding fix as obtained from 5 DF stations.

ceiver of reflected and refracted energy from high points on the horizon. This is particularly true when signals are observed that are either at or below line of sight.

In addition to these sources of error, a lateral deviation on ground wave at frequencies lower than hf is apparent when the wavefront crosses a discontinuity in ground conductivity at other than normal incidence. On other than relatively stable low frequencies, the properties of the propagation paths exhibit a fluctuation and variation with time at different rates, and, during the general observational period, do not have a mean value of zero. Variations in the apparent direction of arrival of a sky wave are also subject to a low periodic component which varies with the sunspot activity cycle.

A second source of error is contributed by irregularities and asymmetry at the antenna array or in its immediate vicinity. Waves may be reflected from nearby objects such as wires, airplane hangers, vehicles, and metallic platforms. Error from these causes are particularly noticeable in the case of airborne direction finders when frequencies are being received at which it is impractical to obtain discriminatory directivity in the airborne antenna.

Irregularities in the conductivity of the ground or image plane close to the DF antenna also produce errors. Poorly bonded connections in a mat or counterpoise can produce errors. A counterpoise or image plane which is too small or irregular in shape will also introduce errors. When striving for a site in which the error is less than 1', it is extremely important to consider many possible sources of error. Buried underground cables, asymmetrical structures beneath the counterpoise, and slight differences in conductivity of the image plane will produce noticeable sources of error. Automobiles parked anywhere near the site will also introduce a disturbance.

The proper location of a direction finder on a warship is one of the most disturbing problems. Reradiation from local reflecting objects may occur with a quadrature polarization. As an example: in the event the direction finder is designed primarily to operate for vertical polarization and a horizontally polarized signal is encountered, the signal energy from a local reflecting object with converted vertical polarization may often be greater than, or certainly of the same order as, the direct horizontally polarized

signal. Since the antenna patterns may be entirely different for each polarization, substantial errors are introduced. This phenomenon is often encountered in airborne antennas, particularly when the antenna is near the shadow of a wing or other portions of the aircraft.

Another source of error is that inherent in the instrument itself. This is generally caused by improper design or faulty operation. In most cases, instrument errors arise from poorly designed antenna elements or goniometers. Factors such as mutual impedance between antenna elements, which is a function of the direction of arrival, produce the so-called quadrantal error in the case of the Adcock. Temperature effects in cables may also be a source of error as may also discontinuities introduced by such components as terminal plugs and connectors. At vhf and uhf, for example, asymmetrical patterns can be produced by the heating of one feed cable by the sun's rays. Instrument errors can also be introduced by indicator nonlinearities and noise sources synchronous with respect to the goniometer or indicator scanning cycle.

In general, the errors from all three of the above-mentioned sources— propagation, site, and instrument—can be classified into two general categories: those which are determinate and therefore may be calibrated out, and those which are indeterminate or of a random nature and are too complex to determine discretely. Instrument errors of a determinate variety can be calibrated by the use of a target transmitter. The target transmitter is moved about on a circumference of a circle four or five wavelengths away from the center of radiation of the structure. When the frequency involved is so low that it is impractical to operate the target transmitter at that distance from the center, the approximate error that may be introduced into the calibrating procedure is obtainable from the curve of Figure 10-82.

To reduce the errors due to path deviations, observational errors, and other phenomena which cannot be calibrated out and are of an indeterminate variety, a number of possible attacks have been used. On the basis that during the observational period certain errors will tend to have a mean of zero, bearings may be taken rapidly over as long a period of time as possible and the average value then determined. In this procedure, computers may be used to automatically indicate the average value.

Another method exists of improving the accuracy of a bearing or of determining which of a series of bearings on the same station are most likely the correct ones. In this method, the operator introduces weighting factors, determined by such considerations as the sharpness or quality of the null in the case of an aural presentation or the approach that a cathode-ray indicator display would have to the ideal case. The gradings are often grouped into one of three classes: either A, B, or C. A class A bearing is a low-error reli-

able bearing, a class B bearing is a usable but inaccurate bearing; and a class C bearing is considered unreliable.
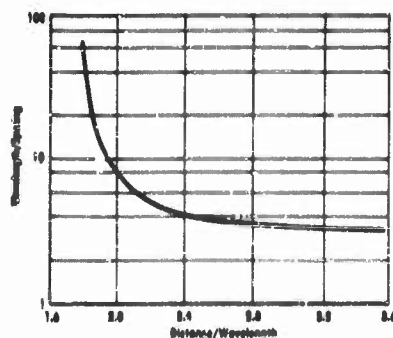
In determining the most probable position of a signal source from two or more bearings at different azimuthal angles to the transmitter, one can concern himself with the deviations which are caused by those different ensemble members which have Gaussian distributions. As indicated previously, in actual practice the statistical errors in a finite integration or smoothing time do not have a zero mean and as a matter of fact do not follow a normal error distribution. Some improvement, however, can be obtained by considering that the errors are of a Gaussian or normal error probability associated with each bearing (References 41, 42, 43).

The error probability of each bearing is measured in terms of its variance, i.e., of the square of the standard deviation of a group of bearings about the true value. The sources of error in the bearing are of various nature and may be classified as follows: *instrumental* (due to imperfect calibration or to fluctuation of the polarization of the wave); *site* (due to spurious reflections from the surrounding terrain); *propagational* (due to reflections from the ionosphere); *observational* (due to intrinsic operator and time causes). Estimation of these errors is obtained in part from past experience and in part from experiment. For example, the observational error is computed by taking an actual number (about 10) of bearings all of which are independent of each other and occur within few seconds' distance in time. From these readings one computes the arithmetic mean, the mean-square deviation, and the minimum readable angle. The corresponding variance is obtained in general with the relation

$$v = r^2/n^2$$

where n is the number of readings taken between 5 and 12, and r is the range of the same readings, i.e., the difference between extreme readings.

The variance due to instrumental errors is generally neglected (unless the calibration is known to be defective); the variance due to ionospheric fluctuations is computed from past experience. For example, in Figure 10-83 a graph is shown which gives the variance for bearings averaged over 5 minutes.

FIG. 10-83. Variance due to ionospheric lateral deviation for bearings averaged over about 5 minutes.

The variance due to site error is also evaluated in advance. As an example: the variance may be about 2° in the bandwidth 3 to 5 Mc, 1° in the bandwidth 5 to 10 Mc, and 0.5° in the bandwidth 10 to 15 Mc.

The total variance corresponding to each bearing is evaluated as the sum of the individual variance as above described. Now, in order to determine the point of maximum probability of intersection, one starts making the few guesses for the position of a point in the area. To correspond to each point P, one computes the quantity,

$$ S = \frac{d_1{}^2}{V_1} + \frac{d_2{}^2}{V_2} + \ldots + \frac{d_n{}^2}{V_n} $$

which is the sum of the ratios of the squares of the angular deviations d_1, d_2, d_3, ... from each of the observed bearings and the corresponding variance of the bearings. The desired point of maximum probability is that for which the quantity S is a minimum. If one computes S for a few points, one immediately is able to improve the accuracy of successive guesses and rapidly arrives at a good solution. In general, a few trials are sufficient for an experienced operator. It may be noted also that the loci of constant difference $S - S_{min}$ are ellipses, and that the ellipses individuate areas in which the probability of finding the point is about the same. For example, if $S - S_{min}$ is 1.4, the corresponding probability of the ellipse is 50%; and if $S - S_{min} = 4.6$, the corresponding probability is 90%.

The application of the statistical method for the determination of the most probable fix has proved very advantageous. Efforts have been made to realize automatic plotters which derive the most probable solution electrically from a set of independent bearings.

REFERENCES

1. Mitra. *The Upper Atmosphere*, Appendix 3, 2nd ed., Asiatic Society, Calcutta.

2. *Reference Data for Radio Engineers*, 4th ed., International Telephone and Telegraph Corporation, p. 667.

3. Mitra et al., *Nature*, 1934, Vol. 133, p. 533.

4. Taylor and Young, *Proc. IRE*, 1928, Vol. 16, p. 561, and 1929, Vol. 17.

5. Kerr, Donald Edwin, *Propagation of Short Radio Waves*, McGraw-Hill, New York, 1951.

6. Grünberg, G. A., "Theory of the Coastal Refraction of Electromagnetic Waves," *J. Phys. (USSR)*, 1942, Vol. 6, p. 185, and 1943, Vol. 7 (errata).

7. Grünberg, G. A., "Suggestions for the Theory of the Coastal Diffraction," *Phys. Rev.*, 1943, Vol. 63, p. 185.

8. Ross, W., and E. N. Bramley, "Lateral Deviation of Radio Waves at Sunrise," *Nature*, 1947, Vol. 159, p. 132.

9. Burtt, G. J., and R. G. P. Whipple, "M.F.D.F. in H. M. Ships," part 2, *J. Inst. Elec. Engrs.*, Vol. 94, part IIIA, No. 15, 1947.

10. "IRE Standards on Navigation Aids· Direction Finder Measurements 1959," *Proc. IRE*, August 1959, Vol. 47, No. 8, p. 1365.

11. Libby, L. L., "Special Aspects of Balanced Shielded Loops," *Proc. IRE*, September 1946, Vol. 34, pp. 641-646.

12. Redgment, P. G., W. Strusaynski, and G. J. Philips, "An Analysis of the Performance of Multi-Element D.F. Systems," *J. Inst. Elec. Engrs.*, 1947, part IIIA, Vol. 94, p. 754.

13. Fletcher, H., "Simple Method of Reducing the Polarization Error," *J. Inst. Elec. Engrs.*, 1947, part IIIA, pp. 771-782.

14. Friis, H. T., and C. B. Feldman, "Multiple Unit Steerable Antenna for S. W. Reception," *Proc. IRE*, July 1937, Vol. 25, No. 7, pp. 841-917.

15. Feldman, C. B., "Deviations of Short Radio Waves from the London-New York Great-Circle Path," *Proc. IRE*, October, 1939, Vol. 27, No. 10, pp. 635-645.

16. deRosa, L. A., "Stochastic Processes as Applied to Aerial Navigation and Direction Finding," IRE National Convention, New York, 1950.

17. deRosa, L. A., "Contribution of Information Theory to the Airborne Reconnaissance Problem," USAF Antenna Symposium (classified), University of Illinois, Urbana, Ill., October 1950.

18. Budde, W. Z., *Nature*, 1950, Vol. 5a, p. 683.

19. Bramley, E. N. and W. Ross, *Proc. Roy. Soc.*, 1951, Vol. A 207, p. 251.

20. Whale, H. A. and W. J. Ross, *Proc. P   r. Soc. London*, 1956, Vol. B69, p. 311.

21. Titheridge, Thesis for M. S., 1956, Auckland University, Australia.

22. Whale, H. A., *J. Atmospheric and Terrest. Phys.*, Vol. 13, No. 34, p. 258.

23. Ross, W., "Fundamental Problems in Radio D.F. at H.F.," *J. Inst. Elec. Engrs.*, 1947, Vol. 94, part IIIA, p. 156.

24. Hatch, J. F., "Development in H.F. D.F.," *J. Inst. Elec. Engrs.*, 1947, Vol. 94, part IIIA, p. 691.

25. Luneberg, R. K., *The Mathematical Theory of Optics*, Brown University Press, Providence, R. I.

26. Earp, C. W. and R. M. Godfrey, "Radio D.F. by Cyclical Differential Measurement of Phase," *J. Inst. Elec. Engrs.*, 1947, Vol. 94, part IIIA, p. 705.

27. Boulet, J. L. L., *Investigation of Doppler Effect in Determining Direction of Arrival of Radio Waves*, Master's Thesis, University of Illinois, Urbana, Ill., September 1947.

28. Boulet, Anderson, and Meara, *Doppler Type D.F.*, Report Number 3, University of Illinois, Urbana, Ill., ONR Contract N6-ORI-71.

29. Hansel, Paul G., U. S. Patent No. 2, 449,553, September 1948.

30. Fantoni, J. A. and R. C. Benoit, "Applying the Doppler Effect to D F. Design," *Electronic Ind.*, January 1957, Vol. 16, No. 2, and February 1957, Vol. 16, No. 1.

31. Baum, Richard, "Design of Directive Broad-Band Antennas," *Proc. IRE*, 1946, Vol. 34, No. 12, pp. 956-959.

32. Sandretto, P. C., *Electronic Avigation Engineering*, International Telephone and Telegraph Corporation, 1958, p. 115.

33. Bendat, J., *Principles and Applications of Random Noise Theory*, Wiley, New York, 1958, p. 370 et seq.

34. Rice, S. O., "Properties of Sine Wave Plus Random Noise," *Bell System Tech. J.*, January 1948, Vol. 27, pp. 109-157.

35. Watson-Watt, R. A. and J. F. Herd, "An Instantaneous Direct-Reading Radio Goniometer," *J. Inst. Elec. Engrs.*, 1926, Vol. 64, p. 611.

36. Wagstaffe, C. F. A., British Patent No. 496,239.

37. Earp, C. W., British Patent No. 490,940.

38. Busignies, H. G., "Un Nouveau radio-compass i'onde electrique," 1930, Vol. 9, p. 397.

39. Busignies, H. G., "Automatic Radio Compass and its Application to Aerial Navigation," *Electrical Communication*, (Tech. J. International Telephone and Telegraph Corp.), October 1936, Vol. 15, pp. 157-172.

40. Stevens and Davis, *Hearing*, Wiley, New York, 1938.

41. Ross, W., "The Estimation of the Probable Accuracy of High Frequency Radio Direction Finding Bearings," *J. Inst. Elec. Engrs.*, 1947, Vol. 94, part IIIA, p. 722.

42. Stanfield, R. G., "Statistical Theory of Direction Finder Fixing," *J. Inst. Elec. Engrs.*, 1947, Vol. 94, part IIIA, p. 762.

43. Barfield, R. H., "Statistical Plotting Methods for Radio Direction Finding," *J. Inst. Elec. Engrs.*, 1947, part IIIA, p. 673.

11

The Analysis of Reconnaissance Information
from the Data Handling Point of View

L. A. deROSA,   E. G. FUE. "T,   J H. VOGELMAN

11.1 Basic Concepts

11.1.1 Purpose of Analysis

The most general purpose of the analysis of electronic reconnaissance data is to develop technical descriptions and to geographically locate the various emitters.

Dependent upon the degree of precision and sophistication of such reconnaissance, the labels of Electronic Order-of-Battle (EOB) or Electronic Intelligence (ELINT) may be assigned. Location information, coarse technical information, determination of functions, and assessment of vulnerability are typical ELINT objectives. The composite of such individual data will generally yield additional intelligence, but the primary function is to compile what may be termed catalog and operational data.

A second general purpose of electronic reconnaissance is to monitor emissions for their semantic aspect (that is, their meaning or message content). This implies a prior knowledge or discovery of the technical characteristics of the signals through ELINT processes. The results of such monitoring are most often called Communications Intelligence (COMINT) when communications messages, as such, are intercepted. However, the concept of semantics can be extended to noncommunications areas: e.g., various guidance or remote control transmissions (where we may extend the concept of semantics

to that of machines communicating with one another). In all such areas, data handling processes are employed not for catalog data generation or comparison, but for message derivation.

These two types of reconnaissance operations may be compared with more conventional communications practice. Reconnaissance may be seen to encompass the problem of completing a communications link with an uncooperative network of transmitters. It is the uncooperative nature of the link between emitters and receiver which sets reconnaissance techniques apart from the more conventional communications art. In conventional communications, for example, one may have prior knowledge of the carrier frequency and modulation type of a set of transmitters. With such knowledge, one may employ a receiver which includes correlation processes (e.g., tuned resonant circuits) in its intrinsic design, and proceed directly to message reception.

In EOB or ELINT practice, one needs to acquire or verify by reconnaissance the same type of information which is known, a *priori*, to a cooperative receiver. This is a task which taxes the receiver and data processing arts even where no message is to be monitored. In COMINT practice, additional tasks are required which approximate or exceed those in cooperative communications, depending on the degree of demodulation or decoding which may be required subsequent to ELINT-level processing. It is evident, therefore, that quite different forms of data processing or signal analysis are required for EOB-ELINT reconnaissance and for COMINT message extraction.

In EOB or ELINT reconnaissance, data processing is employed to generate or verify information which can be described as catalog data. Such data are essentially stationary or redundant in that they are slowly varying with time. The "time-constants" which apply are generally the periods required for the movement of emitters, for logistic schedules, etc. These essentially stationary data are processed by comprehensive and intricate, but familiar, methods of sorting, computation, and correlation.

Sorting is required because the basic data are not properly grouped in discrete sets. Computation is required: (a) because the parameters in which the raw data are received, e.g., direction of arrival, are not those in which the stationary data exist (e.g., position), thus requiring a transformation of parameters; and (b) because statistical processes are required for smoothing and for the generation of quantitative confidence levels. Correlation is required for library-science purposes: to compare the recent data with previous information, thereby preventing redundant library files, providing for updating of those files, and emphasizing "news".

In contrast, COMINT data, in its most general form, is by definition not redundant. Being essentially nonstationary, quite different data processing methods are required. The distinction stems partly from the fact that much

shorter time-constants are associated with the data, that is, the data are changeable and perishable. The fact that the total information, measured in "bits", is normally greater for semantic content than for catalog data contributes to the distinction. Therefore, COMINT data processing is necessarily more subtle and more intricate.

11.1.2 General Character of the Analysis Process

The ch- teristics of a radio transmission, such as radio frequency, type of modul. bandwidth, and other attributes, determine the identity of the transmitter. The knowledge that the signal characteristics must fall within certain limits follows from familiarity with the component art. For example, if a transmitter has an r-f carrier of 100 kilocycles, the bandwidth of the signal most probably will be of the order of 20 kilocycles, or less, since efficient antennas having more than a ±10 percent frequency range at L.F. are infeasible. Unknown advances in technology on the part of an enemy will decrease the constraints and increase the problems of interception and analysis. It is usual to design radio reconnaissance equipment to intercept signals whose characteristics are known to exist. Often surprises are produced by unanticipated progress in the enemy's component art.

The analysis problem, therefore, may be regarded as the examination of the attributes of an intercepted signal and the determination that an orderly collection of such attributes approximates that of a known signal source. If the signal attributes, that is, such characteristics as radio frequency, pulse-width, pulse repetition frequency, rotational rate, and polarization, do not conform with those of a known signal source, either the collection and observation i. far " or a new signal source must be postulated having such new characteristi .

A furt' .. distinction must be made to determine whether the intercepted signal represents a known model but with signal characteristics modified by extraneous effects, such as a poorly operating magnetron in a radar transmitter, or a poorly aligned intercept receiver. Other examples of extraneous effects are propagation anomalies, modulations impressed on the signal from such sources as a tumbling satellite, Faraday rotational changes in polarization, and doppler frequency shifts caused by the relative motion of the source and the receiver.

An important consideration is to determine whether the discrepancy between the observed signal characteristics and the catalog model is due to extraneous effects or whet'.. it represents a change in the intrinsic signal characteristics occasioned by a new version of an old model. In the latter case, the catalog of existing models must be updated and expanded to include the new version.

The reconnaissance process may also be concerned with the actual mean-

ing or semantic aspect of communication signals. The COMINT content of a message in the future, what with the multiplexing of communication signals for directing interceptor operations (GCI), radar surveillance operations, navigational functions and missile guidance systems, will be most difficult to separate from ELINT. To this illustration must be added the use of digital and other more complex encoding methods for the transmission of information, so that whether a signal characteristic is functional and produced by a complex code matrix or whether it is semantic becomes difficult to ascertain until the model has been clearly established and cataloged.

The complete description of an electromagnetic emission may be given by a plot of the instantaneous r-f waveform, the spatial coordinates of the source, and a description of the polarization of the wave, all plotted against time. This plot may be compared with a library of plots of all known signals to attempt a recognition.

Difficulties immediately arise. If the signal is a communication transmitter, the *a priori* knowledge of its message is highly improbable, so that a detailed comparison is impossible since it would be a hopeless task to store all possible messages. A number of simplifying assumptions may be made; e.g., it may be assumed that one of a number of known modulations is present. Detector circuits and display can be contrived to identify the modulation, and the identification may be considered complete when an intelligible message is obtained. This message may be in a known or artificial language.

In general, it is desired to determine the r-f center frequency, bandwidth, type of modulation, signal strength, location of the transmitter, the amount and nature of traffic, and any other characteristics of the demodulated signal instead of a time plot of the r-f wave.
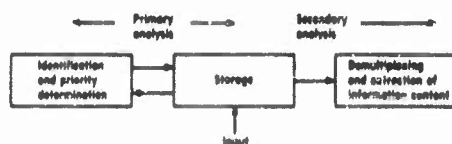
FIGURE 11-1   Functional Diagram
of an Analysis System

A representative functional diagram of an analysis system is shown in Figure 11-1.

In the case of a radar emission, a complete identification would involve a comparison of the time function for at least one antenna rotation (in the case of a surveillance type) with a model from a stored library. This is again a difficult task and in most cases unnecessary. For a radar having an r-f bandwidth of 4 megacycles, negligible information would be required for polarization and spatial coordinates, but with a 1 to 1 signal-to-noise ratio as a usable signal level, approximately 2.5×10^6 bits would have to be stored to analyze every possible evident or hidden characteristic of the system.

This procedure may be necessary for examining the nearby reflections

from the radar for "fingerprinting" purposes, in which case only the time immediately following the main transmitted pulse need be examined. Often many recordings may be superimposed to conserve storage capacity. This is useful when a human observer is available for performing the comparison with the known model.

The human observer is frequently able to identify the patterns obtained and thereby replace much complex electronic equipment. The study of pattern recognition by human observers is a branch of psychophysics. It is possible by suitable transformations to provide patterns which are susceptible to preferential recognition, acoustically and/or visually.

When certain basic assumptions may be made relative to the emission, the analysis function is very much simplified, as exemplified in the following.

The rotational rate of a radar of the surveillance type is generally a constant. Therefore, it is sufficient for the intercept equipment to recognize the rotational rate and unless information is required regarding the antenna directive pattern, the effects of modulation of the r-f signal due to antenna rotation can be neglected and not stored. The rotational rate need not be a constant if it can be expressed by a simple law, so that effects due to the motion of the antenna can thereafter be neglected. A similar analysis can be performed to recognize a nutating-type radar. In addition, if the radar is a pulsed type with a fixed pulsewidth and pulse repetition rate, then further simplifications may be made by recognizing modulations of the r-f carrier as being due to pulses occurring at a certain rate and having a certain width. If the radio frequency of a radar is fixed and does not shift, then the recording of the r-f wave can be dispensed with altogether and a single attribute, radio frequency, may be substituted. By a few simplifying assumptions such as these, the storage capacity necessary to represent the radar signal is reduced from many millions of bits to about one hundred.

In general, information which leads to the identification of a known type of radar, its location, and the number of radars in the community is called radar order-of-battle (ROB) information, whereas detailed information which often is obtainable only by careful analysis of wideband recordings is called technical intelligence. It is much simpler to identify a rad as being one of a known kind for which a model is available than to ana. a complex signal to determine its function and mode of operation.

The simplifying assumptions are often based on fundamental considerations. For example, in procuring a quantity of a certain type of radar for defense purposes, three to six years are generally required to develop an equipment, subject it to environmental tests, and to produce a quantity at a minimum of cost by production line techniques. As is well known, any

change of design during the course of manufacture results in greatly in-
creased costs and delays in procurement and the service and maintenance
problem in the field becomes quite complicated. Because of these conditions,
many radars will have quite similar characteristics and one model may be
used to identify many hundreds of radars. Additional constraints in the
design are introduced by the state of scientific development and component
characteristics. For example, an S-band radar of a given antenna rotation
rate will, within limits, have its beamwidth prescribed, since for a given range
capability (which determines the pulse repetition rate to avoid second time
around echoes) the antenna beamwidth will determine the minimum number
of hits per revolution that will be made on any target. Constraints of this
type dictate that the designer is not free to choose with complete freedom
the characteristics to be designed into a radar emitter.

An interesting method of representing the signal characteristics of an
emitter such as a radar is to plot the simplified attributes—that is, radio
frequency, rotational array, bandwidth, pulse repetition rate, pulsewidth,
and polarization in an n-dimensioned hyperspace coordinate system. A given
radar can then be represented as an n-dimensioned polytope in this hyper-
space. The bounds of such a polytope are not sharp or uniquely determined,
but have a dispersion determined by the tolerances that are maintained
under operating conditions. It is easily seen that not all geometric configura-
tions are possible, and that the constraints placed upon the designer limit
the choice of the geometry of these n-dimensioned volumes.

The problem now is to intersect this polytope with a hyperplane selected
so that a single or a minimum number of measurements will yield the highest
probability of identifying the remaining geometric configuration. This mea-
surement need not be one of a single signal attribute, but rather may be
selected as a relation between two or more attributes. For example, $\psi = \triangle F/F$, where $\triangle F$ is the bandwidth and F is the radio frequency of the
transmitter, may be used. This use of relationships corresponds to an analysis
from the morphological viewpoint and involves the ratio of two edges of the
polytope.

As indicated above, the development of components, practical mechan-
ical design problems and other considerations limit the freedom of choice for
the radar so that the polytope geometry as a function of spatial location has
certain most probable shapes. This means that radars can be identified quite
often by similarities in shape (isomorphism) of their plotted characteristics.
An interesting extension of this is the identification by a comparison of the
Schlegel diagrams* (plane projection methods of n-dimensioned geometry).

*Sommerville—*Introduction to Geometry of n-Dimensions*—Dover, New York

11.1.8 Classification of System Types

The analysis of intercepted signals starts at the receiver. Two distinct philosophies of receiver design have been followed in the construction of reconnaissance equipment—the scanning receiver and the "wide-open" receiver—each of which may be classified according to the type of analysis performed during the collection process.

The scanning receiver selects signals sequentially in order of increasing or decreasing frequency by tuning across the frequency band of interest. The AN/APR-9 is a typical example of such a system. This receiver utilizes a 20-megacycle wide bandpass which can be scanned slowly over a frequency range of 1,000 to 10,750 megacycles. The sensitivity of the AN/APR-9 receiver is approximately 70 decibels below a milliwatt (−dbm), which is usually adequate for detecting radar-type emitters even when the main beam is directed away from the reconnaissance receiver. The major disadvantage of a sequential or frequency scanning system is that in a crowded environment the reconnaissance vehicle may pass through an area of activity before the frequency scanning process has covered the entire band of interest. Also, the probability of missing short messages, such as might be used for GCI operations, is excessively high. A present approach is to increase the probability of interception of short messages by utilizing combinations of scanning receivers and dividing up the responsibilities of signal presence detection and message content detection.

The wide-open receiver technique obviates the previous objection for ELINT application by employing video detection without selection of a narrow-frequency band. The filters and amplifiers preceding the video detector pass the entire band of interest. All signals occurring in the band are processed in parallel by the reconnaissance receiver. The AN/APD-4 and AN/DLD-1 are typical wide-open airborne reconnaissance systems covering the 1,000 to 33,000-megacycle region. Since the bandwidth is wide, receiver sensitivity is only of the order of −40 dbm. Therefore, a receiver with this parallel processing ability may miss signals from distant transmitters located to the side of the vehicle path because of low sensitivity. Another drawback of the parallel processing procedure is the difficulty inherent in processing overlapping signals. For pulsed radars, the overlap of pulses is of negligible consequence; however, the interlace of several pulse trains does pose problems in subsequent analysis. Signals of a high duty cycle, such as c-w radars and most communication signals, cannot be satisfactorily handled by a parallel system.

The parallel or wide-open system is most useful for general area reconnaissance where the electromagnetic environment is defined in broad terms over wide geographical areas. The sequential or "discriminating" system, on

the other hand, is most useful for special area reconnaissance where *a priori* data on the environment are available. In many cases, selective techniques are used as a follow-up to wide-open techniques in obtaining technical intelligence.

11.1.4 Scanning Processes and Probability of Intercept

The concept of "Probability of Intercept" has created much misunderstanding because several definitions may logically be applied. These definitions are based on the probability of obtaining various degrees of intelligence from the intercepted data, which ultimately determines the usefulness of the reconnaissance mission to a particular using agency.

For purposes of comparison, we may select a rather unrealizable but all-inclusive definition that a perfect reconnaissance system must intercept and record* all signals that are emitted at any time, anywhere on earth, over the complete radiation spectrum. However, in obedience to some natural laws and the state-of-the-art, it is reasonable to allow a few compromises in the definition. First of all, let us limit the spectrum to the frequency band usually covered in a' borne instrumentation: that is, 30 megacycles to 40 kilomegacycles. Another reasonable compromise may be to limit the area of coverage to a radius limited by the line-of-sight. Now we should consider the effect of further restrictions in setting a useful criterion, since instead of physical limitations we are now dealing with compromises between the ingenuity of system designers and the needs of intelligence.

If we are capable of intercepting* continuously all signals emitted between 30 megacycles and 40 kilomegacycles at less than line-of-sight from a moving airborne platform, how much of these data are actually needed? If all data are needed, then we must say that this criterion is then our basis for 100 percent probability of intercept.

Considering ELINT alone, the various categories of reconnaissance system designs can usually claim 100 percent probability of intercept under various qualified conditions, but which have varying degrees of usefulness to the intelligence analyst. Accordingly, the following list of categories has been found useful in depicting the amount of intelligence that is ideally available from a given operation. Thus, a system may have 100 percent probability of intercept that can receive, record, and analyze:

(1) All emissions from all radars continuously, in the area covered by line-of-sight regardless of antenna scan, or

(2) All intercepts from main beams of all radars in the line-of-sight area, or

*The assumption is made that all data are so recorded that they can in fact be extracted in the subsequent processing functions.

(3) Sufficient intercepts from main beams in the line-of-sight area for identification and location of all radars.

In this classification order, there is decreasing intelligence capability in the corresponding systems. Thus, a class (1) system can present the full amount of available information regarding any radar that is operated, no matter in what direction its antenna is pointed or for what interval of time. Such a system would obviously be directed mainly toward the area of technical intelligence.

A class (2) system, which utilizes the time scanning of the intercepts of rotating radars to reduce the traffic density is also useful in technical intelligence because there is available a complete time history of all intercepts. The use of reduced sensitivity implies, however, that distant, low-power radars or radars which never "look" at the receiver will not be received.

The class (3) system relies on the abundance of redundancy in a radar environment to reduce both data rate and the required storage capacity. Such redundancy is usually available from scanning search radars, but for short-term trackers and experimental devices there would be less probability of intercept. At present, the class (3) concept appears to be most valuable in reconnaissance for tactical ROB situations.

There may be other probabilities of intercept for special purposes. For example, consider the wide-open type of ELINT reconnaissance receiver in which the operation is typified by a continuous succession of frame or look intervals. The observations obtained during each sampling interval are sequentially recorded on photographic film or magnetic tape. It is pertinent to ask, what is the probability that only one emitter shall be logged within a given frame and in a given environment of scanning radars; or, for a given environment, what is the most desirable frame rate from the standpoint of obtaining a high yield of unambiguous (i.e., single emitter) frames? The Poisson distribution applies:

$$P_r = \frac{a^r e^{-at}}{r!}$$

in which P_r is the probability of exactly r scanning radars logging within a frame of period t in a signal density environment in which the long term average number of radars logged per frame is a.

At this point, it should be mentioned that in general a reconnaissance receiver is designed to scan in one but no more than one of the several available dimensions. If more than one simultaneous scanning process is involved, then the probability of intercept as defined by class (2) or (3) above becomes excessively small. It has been pointed out that if scanning

or preselection in some form is not provided, then the data rate becomes prohibitively high in terms of the required storage capacity. Note that the so-called wide-open ELINT receivers usually entail a tacit scanning process by reason of their limited sensitivity in conjunction with the rotating narrow-beam antennas of the radar emitters. COMINT receivers, on the other hand, are always of the high-sensitivity, high-selectivity frequency scanning variety.

Therefore, for purposes of providing reconnaissance intelligence the problem is that of deciding how much information a system must absorb on a given mission so that the desired intelligence level may be achieved. Since the "rate of absorption" is related to the storage capacity, sensitivity, and frequency coverage of a system, the design is usually a composite of compromises which recognize a limited storage capacity and tends to make use of time or frequency scanning to reduce the data rate.

11.2 Analysis

The problem of analysis of reconnaissance data can be summarized as the *perception of form* in accordance with some form or model already established in fact or concept. Following this definition, this section outlines a conceptual basis for an orderly analysis and generation of end intelligence products. It is to be emphasized that there is no *one* intelligence product, but that there are many. Each user of intelligence produces his own unique requirements and these must be reflected through not only the analysis of data but the entire reconnaissance complex. This involves optimization of collection, storage, processing analysis, and dissemination. In this discussion on analysis, the topic will be divided into two general headings: Primary Analysis and Secondary Analysis.

The primary analysis of reconnaissance data may occur at any point in the system. Indeed, it may take place within the reconnaissance receiver itself. At times it may be difficult to make a clear distinction between the receiving function and the analysis function. Such difficulty may, however, be removed if the primary analysis is thought of as the *utilization* of certain technical parameters generated within the receiver.

11.2.1 Primary Analysis

One of the main difficulties involved in analysis of reconnaissance data is the need for making rather sophisticated decisions at a high rate. The basis of primary analysis is to put first things first. Reference to Figure 11-1 indicates the over-all concept. The data are received from the collection station or stations and put into a storage facility. This facility is in essence an information buffer. Primary analysis operates at a very high speed. The patterns are examined, and identification and priority determination are

performed. For certain intelligence requirements, end products are generated at this point. Basically, the information available may be divided into several categories. For this purpose, we may list the various types of reconnaissance data (as defined in previous sections):

ELINT—intelligence derived *about* the emitter

COMINT—intelligence derived *from* the emitter

RADINT—intelligence derived about and from enemy activities utilizing *active* techniques

With these definitions in mind, we may broadly list the information which can be derived during the primary analysis. From ELINT data we can determine the type or category of emitter (e.g., Token, Yo Yo, etc.). For radars this requires sorting and correlation of amplitude-modulation time series. In addition, we may determine the level of emitter activity or average density of any number of parameters, such as frequency-time, frequency-position, prf-frequency, prf-pulsewidth, etc. From communications data we may determine the classification of the emitter, e.g., ground-to-ground communications utilizing hand-set Morse, frequency, stability, location, priority, traffic analysis, etc. In the case of COMINT we can determine the intelligence that is being sent over the circuit. However, since this requires a far more detailed and time-consuming analysis, it is *not* performed as part of preliminary analysis. This function is performed as part of secondary analysis, but only after the priority determination has placed first things first.
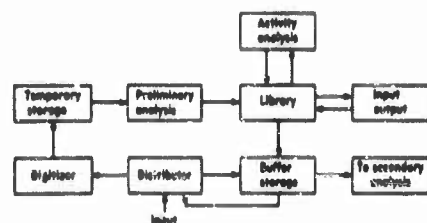
FIGURE 11-2   Primary Analysis

A more detailed picture of the primary analysis concept is given in Figure 11-2. The distributor routes all intercepts to preliminary analysis and to the buffer storage. If the input data are such that primary analysis becomes over-loaded, then data entry from the buffer may be effected at a later time. In brief, the primary analysis operation consists of digitizing the signal for a preliminary analysis. Preliminary analysis extracts video parameter data necessary for signal identification. Actual identification is performed by a table-lookup operation in the library which is the heart of the primary analysis operation.

11.2.2 Library and Indexing

The data extracted from the library are also utilized to obtain an activity

analysis as mentioned before, e.g., density counts, traffic analysis, etc. In addition, the extracted data is used for priority determination. Once the priority has been established, it is assigned to the signal that is stored in its entirety and true form in the buffer storage. In this manner, updated priority assignments are available to the buffer to control the appropriate sequencing of input data to secondary analysis.

It is clear that the success of the preceding method of primary analysis depends upon two factors: generation of an indexing word and library indexing of *a priori* data. The heart of the analysis at this stage can be simply stated. The input signal must be characterized by an uniquely descriptive indexing word. This word must be digital to be compatible with the high-speed search and logical comparison operation of the table-lookup function required of the library. The system design must so devise the signal indexing word that a rapid, accurate, reliable comparison operation can be performed. The indexing operation consists of two steps: signal analysis and merger
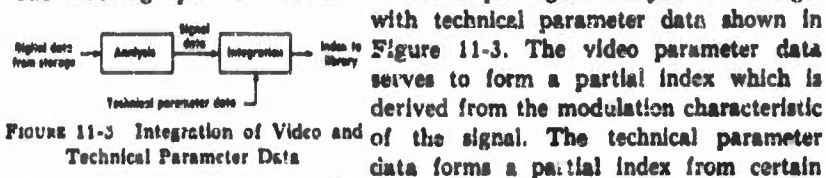
FIGURE 11-3 Integration of Video and Technical Parameter Data

with technical parameter data shown in Figure 11-3. The video parameter data serves to form a partial index which is derived from the modulation characteristic of the signal. The technical parameter data forms a partial index from certain

data generated within the collection system. These are items such as frequency, time, location, propagation conditions, etc. Merged in the proper manner, they form a digital word that may be used to identify an address in the library. Once this address is located, logical operations may proceed to further subdivide the index and the address as the category and identification processes increase in detail.

It is clear that prior knowledge of the signal category is necessary before library identification techniques can be applied to the problem. This implies use of all existing information regarding signal types and usage. An extension of the library-indexing process is information feedback to the library. In case an indexing word is generated during primary analysis which is not identifiable with any words stored in the library, this is noted and human intervention is required. If this signal appears to be in a legitimate but new category, the proper index is given an address in library storage. In this manner the library may be made to regenerate itself based upon reconnaissance data coming in from collectors or intelligence derived from other sources.

In summary, the function of primary analysis is to permit reliable high-speed indexing, classification, and priority assignment of incoming signals. It permits determination of the optimum order of higher level analyses which

utilize man's deductive and inductive capacities as well as advanced machine techniques.

11.2.3 Secondary Analysis

After initial analysis has generated the first-order intelligence products by establishing the category and priority of the intercepted data, the *original* data are transferred to secondary analysis. Since secondary analysis is more detailed, it proceeds at a lower rate. The basis for determining the sequential order of signal processing is that part of the indexing word which denotes priority. A full discussion of priority is given elsewhere. Briefly, it deals with the dynamic behavior of the "value" of the intelligence. This implies a time-value relationship. This functional relation does not itself remain invariant but is a function, in turn, of other factors: e.g., pre possibilities, revision of validation function, etc.

The purpose of secondary analysis is to determine detailed information concerning the signal to accomplish two things. First: in the case of non-communications signals, to permit a *unique* identification. Second: in the case of communications signals, to permit not only unique equipment identity but to obtain message content and, in the case of hand-sent signals, operator identification.

The success of both of these objectives depends upon a close scrutiny of the signal over a long time interval. Thus, continuity of intercept is a fundamental requisite. This should not, however, rule out a sampling approach *per se*, since such a transform might be far better adapted to a particular kind of intercepted signal and may carry the desired amount of information.

FIGURE 11-4   Secondary Analysis

Figure 11-4 shows the secondary analysis concept. In order that a carrier containing *n* independent channels be handled, it is first necessary to recover the base band or bands and then supply appropriate demodulation for each channel. Automatic, flexible machine techniques have been developed that permit such a demultiplexing operation even for very complicated systems.

A character analysis is performed on the data to determine the nature of the transmission, e.g., machine Morse, teletype, etc. After the mode of emission has been determined, it is generally possible to convert all pulse codes to a common machine language. In the case of non-pulse transmissions, it is

necessary to apply sampling techniques and convert to pulses or to bring in a human analyst at this point. A decision regarding which course of action is best must rest with an analysis of all factors in a particular case. It is dangerous to generalize on this point.

Assuming that a common machine language has been achieved, communications signals are examined to determine the linguistic origin. A number of methods of approach are open here, e.g., keyword analysis, phonemic frequency distribution, etc. If the language base can be determined, translation will probably be required. At the present time, there are no completely satisfactory machine translation techniques. Hence, a human may be required at this point. It is probable, however, that satisfactory automatic methods will become feasible in the near future.

If translation is impossible by any means, it is reasonable to assume that the data is encrypted. A cryptographic analysis will reveal this fact as well as the level of crypt. This is useful for determining additional priority information as well as aiding in actual decryption.

In the case of message content, a complete text analysis will reveal much detailed information of interest. If the communication follows a known standard operating procedure, much of the text analysis can be machine implemented. However, there will be many instances where it will be more effective to utilize human capabilities. For example, it is not presently possible to instrument sophisticated inferential techniques—particularly where decisions regarding various "shades of meaning" are involved. This is still a fruitful area of human endeavor.

Finally, the dissemination of the intelligence information must be given careful consideration. This area involves transformation of information into a meaningful, timely product that is compatible with user requirements. As such, approximate communications channels, display devices, information call and recall capabilities must be provided. The dissemination mechanism is very diffuse because it involves access to nearly all parameters discussed thus far, e.g., density, index data classification, identification, alarms, etc.

11.3 Techniques and Devices

11.3.1 Storage Systems

A major common requirement of all reconnaissance systems is that of data storage capability. This requirement arises from several causes, such as: (a) a large amount of library type information is required in order to ascertain the significance of new information, (b) a large buffer storage capability is required to enable the processing of high priority data without loss of data of lesser priority, (c) it is usually difficult—if not impractical—to com-

pletely process reconnaissance data in real time, as, for example, when data are being gathered by a reconnaissance drone, and (d) the analysis process itself entails a variety of preliminary storage applications, such as providing a means for determining the pulse repetition rate of a radar type emission.

All of the data storage techniques and devices that have been applied to computer technology have also been used to implement the processes of collection and analysis of electronic reconnaissance data. The media wi ·b predominate are magnetic tape, photographic film, punched cards, magnetic cores, electromagnetic ·lay lines, sonic delay lines, cathode ray storage tubes, printed copy,   encil and paper.

There is much room for growth in the realm of storage systems. It would be highly desirable, for example, to provide for long term storage of portions of the r-f spectrum above 30 megacycles. Such a facility would greatly increase the probability of detection of transmissions utilizing techniques or systems of which we may be unaware, now or in the future, simply because existing detection systems are not sensitive to all possible classes and modes of conveying information via electromagnetic energy.

11.3.2 Analysis Techniques

There is a tremendous range of degree of sophistication of techniques invoked for the derivation of ELINT and COMINT from reconnaissance missions. This is to be expected, since the information derived ranges in complexity from deriving the radio frequency of an emission to determining the total significance of an encrypted foreign language message, taking into consideration any pertinent information gathered from other sources. No attempt will be made to cover all techniques that have been utilized, as such an endeavor would be well beyond the scope of this book. Rather, we shall attempt to cor   a general approach by means of a few specific examples.

Long term storage devices are as yet inadequate for the radio frequencies customarily used for radar and military communications; therefore, the frequency of an emission is always determined in some fashion, coded, and then stored. The method of frequency measurement used in conjunction with the frequency scanning type of receiver is obvious, but some of the techniques applied to the wide-open type of receiver are worthy of mention. Instantaneous frequency measurement is often accomplished by using a bank of overlapping bandpass filters driven in parallel from a single antenna, electronically (or otherwise) observing, interpolating and finally recording an analog of the actual radio frequency. Another technique, used in the AN/APD-4, depends upon the relationship between the directivity of the antenna array and the radio frequency of the received emission.

The pulse repetition frequency of a radar type emitter may be determined by measuring the burst length, counting the number of pulses comprising the burst, and dividing the latter by the former. Or, pulse repetition interval (PRI) may be measured by starting a clock-driven counter with one intercepted radar pulse, stopping the counter with the following radar pulse, and reading the contents of the counter. If the clock is operating at a 1-megacycle rate, the reading of PRI is in microseconds. It is customary to check that the two pulses emanated from the same radar emitter by comparing the time interval between the first and second received pulses with the time interval between the second and third received pulses. If the two intervals are not equal or nearly equal then the three pulses may generally be presumed to have emanated from more than one radar.

The width of a radar pulse may be determined by using differentiation or integration methods. In the case of a wide pulse, differentiation is preferable. The detected pulse is passed through a differentiating circuit and a counter is employed to measure the time between the differentiated leading and trailing edges. In the case of narrow pulsewidths, integration of the detected pulse is commonly used. The integral or area of the pulse is then an indication of the pulsewidth. A weighting function must be introduced to account for the normalization of the pulse height.

There are several techniques appropriate for indicating signal density as a function of time or geographic location. The CP-269/ASQ-18 system accomplishes this function by plotting frequency of intercept versus time at the system frame rate of 60 points per second. The high-speed plotting is accomplished by using a light-beam galvanometer recording on self-developing, light-sensitive paper. Thus, in a relatively short time a display is generated which reveals the total number of emitters intercepted, the number of active emitters at any given time or location, as well as other valuable data such as the approximate frequencies and rotation rates.

The polarization of an incoming electromagnetic wave is determined by the response to the wave of an antenna having known polarization features. Four antennas, one each vertically polarized, horizontally polarized, right-handed circularly polarized, and left-handed circularly polarized, will determine the polarization of the incoming wave. In many reconnaissance systems it is sufficient to know whether the signal contains either vertical, horizontal, or both components, in which case the two circularly polarized antennas are not necessary.

The rotation rate of a radar antenna is determined by counting the number of main-lobe hits in a given time interval. If the information is in long-term storage on film, then the observer has only to count the number of bursts in a given length of film and, knowing the film speed, determine the

rotational rate. In case the information is in long-term storage on magnetic tape, read-out can be made at the record speed through a peak reading detector to a counter used in the same manner as in pulse repetition frequency determination. The peak reading detector is used to smooth out the fine structure of the radar pulses to yield one pulse per burst.

The operational function of each emitter may generally be learned from observation of the fine grain structure of specific parameters or combinations of parameters. For example, a radar that exhibits a low pulse repetition frequency in combination with a slow scan rate and relatively low (for radar) radio frequency quite apparently is used for early warning.

Emitter location is usually ascertained through application of one or another of the direction-finding techniques enumerated in another chapter. The ASQ-18 system, designed for a portion of the r-f spectrum where direction-finding antennas are incompatible with present-day aircraft, represents an exception in that emitter location is determined by correlation of the variation of recorded signal strength as a function of aircraft headings and geographical location with the known radiation pattern of the receiving antennas. There are other possibilities, as, for example, emitter location by analysis of scan rate doppler.

A portion of COMINT processing may be automatized by the use of decision circuits, such as AM or FM, pulsewidth modulation or pulse code modulation, etc., which automatically route a message through the proper matrix to accomplish appropriate demodulation and demultiplexing. Much more COMINT processing will eventually be accomplished by machine methods, including translation and description, but for the present the specially trained human analyst is by far the most important factor.

11.3.3 Sorting Methods

A frequent major requirement in processing reconnaissance data is that of sorting the data in such a way or ways as to yield the maximum possible amount of intelligence material.

One of the more glowing examples of the requirement for large scale sorting appears during the processing of data gathered by the ASQ-18, which is a wide-open airborne receiver intended for the detection and an 'sis of intercepts emanating from radar type emitters. The receiver has a fr ne rate of 60 per second and the data collected in each frame is recorded on magnetic tape. The CP-269 is used to convert the information on this first magnetic tape from analog to digital form and record it frame by frame on a second magnetic tape in a format compatible with standard IBM data processing equipment. These data frames are recorded in the air and transcribed on the ground in the same order as the aircraft was illuminated by

the many radars comprising the environment through which the aircraft passed. Thus, because of rotation rates, burst length, peculiarities of radiation patterns and varying transmission distances, these data are very disordered as well as voluminous. It is desired to organize the data according to emitter, and so a large scale sorting process is next required.

A special program has been written to accomplish the sorting operation on an IBM Type 704 data processing machine. The method that evolved is somewhat unique because of the large size of the sort.

The magnetic core storage of the 704 is programmed to perform the function of 20,000 individual counters. Each counter is used to tally the number of intercepts occurring within its associated three-dimensional domain of frequency, PRI, and pulsewidth space. The 20,000 counters maybe "arranged" in any desired configuration to encompass whatever portion of frequency, PRI, and pulsewidth is desired and with whatever degree of fineness is deemed most appropriate along each axis. After all intercepts have been tallied by these counters, a search is made for that counter registering the highest tally. The intercepts that registered on this particular counter are labeled Unit Number 1. A search is then made for the counter with the next greatest tally. If the cell it represents shaves a boundary point, line, or face with the cell already labeled, its intercepts are included as a part of Unit Number 1; otherwise its intercepts are labeled Unit Number 2. Similarly, the third highest reading counter (cell) is examined for adjacency with any previously examined counter (cell) and appropriately labeled, and so on until all or most all intercepts have been associated with one of the clusters so formed. The mean and standard deviation is computed for each parameter of each cluster of intercepts. If one or more of a set of readings of standard deviation associated with a particular cluster are commensurate with experience, then the intercepts comprising that cluster are deemed to be from a single emitter and are sorted out and made available for further action. If a standard deviation reading is excessive, then that portion of frequency, PRI, and pulsewidth space is redivided into a finer structure of cells and the process is repeated until the intercepts have all been sorted according to emitters.

* * * * * * *

Part III

Jamming and Deception

12

Basic Types of Masking Jammers

Y. MORITA,   R. A. ROLLIN, Jr.

Previous chapters deal with signal intercept and the problem of deter-
mining the presence and nature of an enemy's sensors and weapons. This
chapter deals only with the general characteristics of masking jammer trans-
mitters (henceforth called masking jammers), jammers capable of obscuring
or denying information.* Denial, which is a characteristic of the effect pro-
duced by masking jammers, is achieved by submerging data in interference
which may consist of white noise or signal-like noises. Since masking jam-
mers may be used against radars, communication receivers, or fuzes, the
nature of the interference used will depend upon which of these is being
jammed. Thus, transmission of a narrow band of noise may suffice to make
speech over a radio link unintelligible to a listener, while transmission of a
wide band of noise may be necessary to deny information to a radar using
a correlation process for detection.

The complexity required of masking jammers also depends upon the
device being jammed. For example, if a jammer is being used against a
frequency diversity radar, receivers and programming equipment may be
required to tune the jammer to the correct frequencies. It may be impossible
to predict or determine in time the next jamming frequency required to
successfully counter a frequency diversity system using nonperiodic fre-
quency control. Since specific jammers can take on many forms, only a few

*Radar repeaters and transponders, the subjects treated in Chapter 15, fulfill the
function of creating false information.

generic types are described in this chapter; some jammers will include features of two or more of the basic types.

12.1 Factors Affecting Masking Jamming

Factors which affect masking jamming include the bandwidth, power, and modulation of the jamming transmitter, the antenna characteristics (Chapter 29), the geometric relation between the receiver being jammed and the jamming transmitter (Chapter 13), and the environmental effects on propagation (Chapter 31). Bandwidth, power, and modulation are discussed in this chapter.

The bandwidth of a jammer and its radiated power are both important factors in masking jamming. These factors are related such that a design change in one often requires a change in the other. For example, a requirement for increasing the already wide bandwidth of a barrage jammer (Section 12.4.1) may arise. It may not be possible to maintain the average power level over the extended frequency band if the components (primarily the output tube) and design of the jammer are not changed. As the ranges of radars and transmitters are increased by means of increased power or increased sophistication, the demand for higher jammer outputs also increases. This demand for increased jammer power plus the desire for the capability to jam selectively has resulted in spot, swept, and swept-lock jammers (Sections 12.4.2, 12.4.3, and 12.4.4). In these masking jammers, the noise power is concentrated in a narrow frequency band; the position of this narrow band within a large frequency band can be adjusted. Tubes which can furnish the necessary high power output within a narrow band are usually more readily available than tubes which can furnish high power output over a wide bandwidth.

Another important factor which affects masking jamming is the type of modulation used. The type of modulation, such as amplitude or frequency modulation, and the modulating waveform, such as a sine wave or a saw-tooth function, determine how the available jamming energy is distributed in the frequency spectrum. It is the energy in the sidebands that ordinarily is effective in producing jamming. Noise was recognized early as having characteristics which would be desirable as a modulating source. Considerable effort during World War II was focused on the analysis of noise signals and their effects on receivers (References 1 and 2). Short verbal descriptions of some noise-modulated signals are given below; the jamming effectiveness of several of these signals is discussed in Chapter 14. (Antijam or countercountermeasure features of radar receivers, features designed to reduce the effectiveness of masking jamming, are also discussed in Chapter 14 and in References 3 and 4.) One type of signal, the direct amplification

of noise (DINA) does not involve any modulation. In this case, the energy
is distributed evenly over the bandwidth; the amplitudes follow a gaussian
distribution. DINA is often used because of its applicability as a jamming
signal against many types of signals.

The spectrum obtained by frequency or phase modulation by random
noise alone is approximately gaussian when the ratio of the upper fre-
quency limit of the noise band used for modulation to the frequency depar-
ture from the carrier (for a modulating voltage equal to the rms of the
noise) is approximately 1 (Figure 12-1a and Reference 5). Such an energy
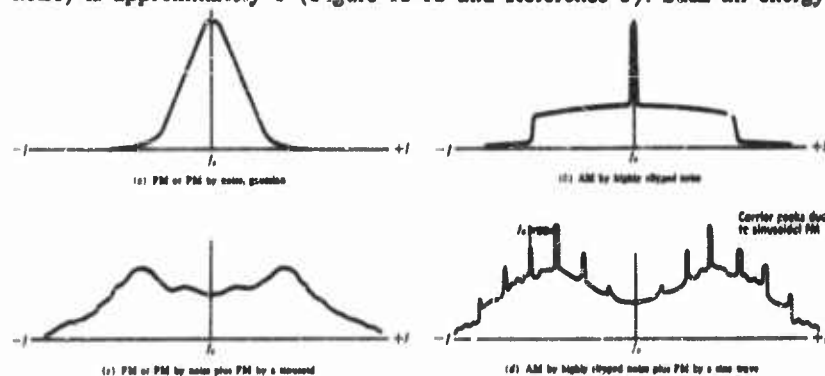
FIGURE 12-1. Noise modulation spectra (References 2 and 3).

distribution is not sufficiently uniform or broad enough to be used with
good effect in barrage jammers. If the ratio is increased to the neighbor-
hood of 3.5 for FM or 2.5 for PM, the energy distribution flattens out and
becomes more uniform. The high ratio indicates that the modulating band-
width is wider than the transmitted bandwidth, and that it might be more
desirable to use direct noise amplification.

When a carrier is amplitude modulated by noise, the amplitude of the
distributed energy is proportional to the amplitude of the modulating noise
and the spectrum is twice as broad as that of the modulating noise. The
spectrum of the carrier can be represented by $C^2\delta (f - 0)$ where C^2 is the
mean square amplitude of the carrier component and $\delta(f - 0)$ is a delta
function with the property of being infinite at $f = 0$ and zero elsewhere.
There are two reasons for not using amplitude modulation: first, consider-
able energy is retained in the carrier; and second, a "ceiling" is placed on
the energy in the sidebands. If the noise is not allowed to be clipped, and
randomness is to be maintained, the modulation level must be kept very
low. In this case, all the energy is in the carrier. When clipping is applied,
more energy appears in the sidebands. Even with extreme clipping, half

the energy remains in the carrier (Section 12.2). When the signal is clipped, the energy spectrum is no longer flat across the sidebands (Figure 12-1b).

It is possible to combine noise modulation with modulation by a non-random, periodic waveform to obtain a jamming signal which can be used over a broad portion of the frequency spectrum. For example, frequency modulation by noise can be combined with frequency modulation by a sine wave. Frequency modulation by sine wave alone produces a spectrum of frequencies located at frequencies $\pm n f_o$ about the "carrier," f_o. (f_o is the frequency of the modulating sine wave.) While it may be broad enough, the spectrum can be sufficiently discrete to provide unsatisfactory barrage jamming. When frequency modulation by sine wave is combined with frequency modulation by noise, a spectral distribution of noise is obtained around each of the sideband frequencies, $n f_o$ (Figure 12-1c). All the energy available for jamming is in the sidebands, a desirable characteristic since the jamming spectrum is broadened and no frequency "holes" appear in the spectrum.

Amplitude modulation by noise may also be combined with frequency modulation by noise to obtain a spectrum suitable for barrage jamming. A broad spectrum is obtained in this case but one which is not as uniform as when frequency modulation by sine wave and frequency modulation by noise is used. Peaks are produced by the various carriers or sinusoidal frequency-modulation sidebands (Figure 12-1d). As in the case of extremely clipped amplitude modulation by noise, energy in the sidebands is limited at best to one-half the total energy.

Signal-like noises can also produce confusion. "Railing" is a technique in which pulses are transmitted asynchronously at the pulse repetition frequency of a radar. A similar technique involves transmitting pulses at random so that the display of a search radar becomes cluttered.

12.2 Clipping

When the effectiveness of noise signals is considered, the distribution of amplitudes is often assumed to be gaussian. In practice, the distribution is not truly gaussian because the high noise peaks are sliced off either intentionally or because of equipment limitations. It has been shown that when using DINA or amplitude modulation by noise, controlled clipping can reduce peak power requirements and, in the latter case, can minimize power wastage in the carrier (Reference 1, pp. 85-88). Controlled clipping permits a more efficient use of the peak power capability of a jammer transmitter's output tube. How effective clipping is depends upon the bandwidth of the receiver against which a jammer is being used. If the receiver bandwidth is about the same as the noise bandwidth, clipping can place a ceiling on

the deflection of a recording device. If the receiver bandwidth is narrow compared to the clipped noise bandwidth, the noise signal will appear to be gaussian to the receiver and will have the same effectiveness as gaussian noise.

12.3 Look Through

Once a jammer has been turned on, some provision must be made to monitor the frequency of the radar or radio jammed so that the jammer can transmit in the proper frequency band. This capability for monitoring the radar or radio signal is called "look through." "Look through" presents a problem because the monitoring receiver must cope with the jammer signal. One way of achieving "look through" is to turn off the jammer periodically. There are two variables which can be changed: the duration of the jammer "off" period and the frequency with which this duration occurs. How these variables are changed depends to a large extent upon the system being jammed. Thus, many looks of short duration might be used against a frequency-diversity radar while a single look might suffice against a fixed frequency radar.

Another way of achieving "look through" is to use correlation techniques. If the received signal can be correlated with a portion of the transmitted signal, that part of the received signal which is due to leakage from the jammer transmitter can be separated from the radar signal and rejected. A coaxial cable or a waveguide can be used to furnish an appropriate portion of the noise signal from the jammer to the correlator. If pseudorandom noise is used, it would be possible to use two synchronized linear-sequence generators, one as a modulating source for the jammer and the other as a source for the correlator. The only connection required between the jammer and the correlator would be a line used to synchronize the two pseudonoise generators; the connection need not be a coaxial cable or a waveguide.

12.4 Types of Jammers

Masking jammers may be placed in the categories of barrage jammers, spot jammers, swept jammers, and sweep-lock jammers. Some jammers fall into two or more of these categories by having alternative modes of operation. Wideband frequency coverage is obtained by many jammers by using multiple transmitters, each covering a given portion of the spectrum. Only the jammer transmitter is discussed in this section. Receivers used for detecting the presence and character of signals are discussed in Chapters 6 and 9.

12.4.1 Barrage Jammers

Barrage jammers are wideband noise transmitters designed to deny use

of frequencies over wide portions of the electromagnetic spectrum; these jammers may be used against radar and communication receivers. The use of this type of jammer is attractive because a number of enemy receivers can be jammed simultaneously or because frequency-diversity radars (or other types of frequency-diversity systems) can be jammed without readjusting the jamming frequency. The type of output tube used determines the effective jamming bandwidths covered by barrage jammers. Typically, at X-band, a magnetron may achieve a 15 percent bandwidth, while a carcinotron (a M-type backward-wave oscillator, Chapter 27) might achieve a 30 percent bandwidth. Generally, the percent bandwidth is wider at lower frequencies; a klystron at S-band may have a 4 percent bandwidth while another klystron at L-band may have an 8 percent bandwidth.

The block diagram of a typical barrage jammer is shown in Figure 12-2.

FIGURE 12-2. Typical barrage jammer.

The modulating signal is amplified noise. The type of modulation determines whether an rf amplifier or an rf oscillator is used for the final stage. If frequency modulation by noise is desired, the last stage might be a voltage-tunable oscillator such as a carcinotron. If the output desired is noise by amplitude modulation, the last stage would be an rf amplifier.

The advantages of barrage jammers are their simplicity and their ability to cover a wide portion of the electromagnetic spectrum. (The barrage jammer was used widely in World War II because of its design simplicity.) The latter advantage can turn into a disadvantage when the systems against which the jammer is working utilize high-powered transmitters. In such cases, the jammer power may not be sufficiently high to effectively mask in the receiver the transmitted signals. The trends in radars to high power and frequency diversity have caused designers to seek means of extending the power and bandwidth of output tubes for barrage jammers.

12.4.2 Spot Jammers

Spot jammers are narrowband, manually tunable transmitters which are amplitude or frequency modulated by random noise or by a periodic function. These jammers are used to mask specific transmitters from communications or radar receivers. Spot jammers can deny range and angle information to radars and can degrade the intelligibility of speech or of other types of modulated signals in communications receivers.

A block diagram of a spot jammer is shown in Figure 12-3. Many spot jammers use amplitude modulation by noise in which at least half the rf energy remains in the carrier. The output power spectrum of a spot jammer
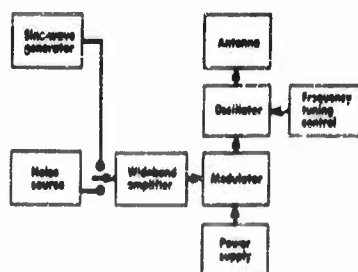
FIGURE 12-3. Typical spot jammer.

can be continuous over a band up to 5 percent of the carrier frequency because oscillators such as magnetrons are frequency modulated by the frequency pushing factor.

Intercept panoramic receivers must be used with these jammers in order to "set" the transmitter frequency on the frequency of the radar or communications sets being jammed. A "look through" capability is desirable so that the jammer frequency may be kept on the frequency of the transmitter being jammed. How accurately a spot jammer must be tuned to a given frequency depends on the bandwidth of the jammer and on the service against which the jammer is being used. If the jammer spectrum is narrow and the service being jammed is keyed cw, the jammer must be tuned to within a few cycles per second. If the jamming spectrum is broad and the service being jammed is radar, tuning to within 0.5 mcs may be sufficient.

The chief advantage of spot jammers is that their output may be concentrated in a narrow power spectrum. Thus, the jammers have the capability of countermeasuring a radar or communication receiver at longer distances than can a broadband or barrage-type jammer of the same power output. Or, for a given distance, a spot jammer can be smaller and lighter than a barrage jammer. Since spot jammers can concentrate large amounts of power in a narrow spectrum, these jammers have the capability under certain conditions of inserting enough power in a receiver to saturate the i-f amplifier and to reduce the gain of this amplifier by its agc action. The conditions include those of short range and proper orientation of antennas. Where space requirements prohibit tunable receivers as an antijam feature (i.e., in VT fuzes and in some missiles), spot jammers can be used to good advantage.

Disadvantages, as well as advantages, arise out of the narrow frequency spectrum of spot jammers. An operator must be available to tune the jammer. The application of spot jamming requires a jamming transmitter for each radar or communication transmission channel to be masked or degraded. The complexity of spot jammers is increased when they must be used against transmitters capable of rapid tuning. Some means of rapidly retuning the jammer to the proper frequency must be provided. Power may be wasted by concentrating too much jamming power in a single channel.

12.4.3 Swept Jammers
Swept jammers are transmitters in which a narrowband jamming signal

can be tuned over a broad frequency band. These jammers have been developed to combine the high power capabilities of spot jammers and the broad bandwidth of barrage jammers. Swept jammers can be employed effectively against radar and communication receivers. The jamming signal is generated in a narrow frequency band, and this band is then swept over a broad portion of the frequency spectrum. Figure 12-4 is a block diagram of a typical tuned swept-jamming transmitter. The output tube is modulated by noise or by a periodic function; frequency or amplitude modulation may be used.

Two factors which are important in swept jamming are the noise power per megacycle and the sweep rate. The sweep rate, the bandwidth of the swept jammer, the bandwidth of the receiver being jammed, the geometric relation between the jammer and the receiver, and the characteristics of the jammer and receiver antennas all play important roles in determining the dwell time, the period during which the jammer noise is in the receiver bandpass. All these factors must be taken into consideration in order to maintain a balance between the dwell time and the periods between dwell times; otherwise, a swept jammer can be ineffective.

FIGURE 12-4. Basic block diagram of swept jammer.

Swept jammers combine the advantage of concentrating noise power in a narrow band and of effectively covering a large bandwidth. Such jammers can be used more effectively than spot jammers against radar nets in which the various radars are tuned to different frequencies. Several swept jammers, sweeping rapidly at different rates, can with high probability, obscure most signals in a given frequency band.

Because of the large number of factors affecting the effectiveness of a swept jammer, there must be comprehensive knowledge of the systems against which the jammer is to be used. The swept jammer will also be more complex than either the spot or barrage jammers.

12.4.4 Sweep Lock-On Jammers

A sweep lock-on jammer is a transmitter in which a narrowband jamming signal can be tuned over a broad frequency band and the signal locked on a particular frequency. This type of jammer is essentially a swept jammer with the additional feature of lock-on capability. A block diagram of a typical sweep lock-on jammer is shown in Figure 12-5. A receiver and the jammer transmitter are swept over the same frequency band. When the receiver
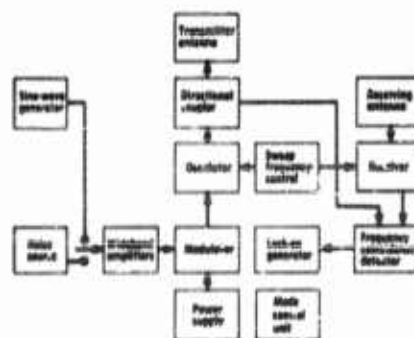
FIGURE 12-5. Basic block diagram of sweep lock-on jammer.

encounters a signal, the frequency sweep is halted, and the jammer transmitter acts as a spot jammer at that frequency. By providing the jammer with a look-through capability, the receiver can be made to start sweeping again when the original signal being jammed disappears. It is well to point out here that many jammers are constructed to operate in several modes, i.e., they are capable of operating in spot, swept, or sweep-lock modes (References 6, 7, 8).

Sweep lock-on jammers can also be programmed in various ways. One way is to sweep the jammer signal and the receiver over a specified band. Another way is to sweep the receiver over a specified band until a signal is received and then to turn on the jammer transmitter at the received frequency.

The sweep lock-on jammer, like the spot jammer, can concentrate much noise power in a narrow band. In addition, it can lock on to a second signal much more quickly than can a spot jammer.

The sweep lock-on jammer suffers from the same limitations as the spot jammer. Only a narrow frequency band can be jammed at any time and more noise energy than is required may be concentrated in that band if the jammer is being used against a receiver other than the ones for which the jammer had been specifically intended. The automatic tuning feature of lock-on jammers can cause two or more such jammers to lock on each other's transmissions. Care must be exercised in assigning frequency channels to each of the jammers, or circuitry must be devised to detect only those signals to be jammed.

REFERENCES

1. *Radio Countermeasures*, Summary Technical Report of Division 15, NDRC, Vol. 1, National Defense Research Committee, Washington, D.C., 1946 (SECRET).

2. D. Middleton, *The Spectra of Noise-Modulated Waves and Their Relative Efficiencies for Barrage Jamming*, Report 29 (ATI 27188), Radio Research Laboratory, Harvard University, Cambridge, Massachusetts, 23 April, 1943 (UNCLASSIFIED).

3. *Anti-Jamming Techniques Study, Final Report*, AD 148671, Sylvania Electric Products, Inc., Applied Research Laboratory, Waltham, Massachusetts, 30 October 1957 (SECRET).

4. *Research on Anti-Jamming Techniques, Final Report*, General Electric Company, Advanced Electronics Center, Ithaca, New York, October 1957 (SECRET).

5. J. L. Stewart, *The Power Spectra of Signals Phase and Frequency Modulated by Gaussian Noise*, Technical Report No. 23, Electronic Defense Group, The University of Michigan, Ann Arbor, Michigan, November 1953 (UNCLASSIFIED).

6. *Final Engineering Report for Carcinotron Jammer QRC-76(T)* HALL-17456-F (AD 308895), The Hallicrafters Company, Chicago 24, Illinois, 18 May 1959 (SECRET).

7. *Flight Test Report for Jamming System QRC-23(T) vs. Radar Tracking Systems AN/MSQ-1A*, WADC Technical Report 57-418 (AD 142083), Weapons Guidance Laboratory, Wright Air Development Command, Wright-Patterson Air Force Base, Ohio, 15 November 1957 (SECRET).

8. *Countermeasures Transmitting Set AN/ALT-6B*, Handbook Service Instructions (Preliminary), General Electric Company, Utica, New York (CONFIDENTIAL).

13

Geometry of the Jamming Problem

Y. MORITA,   D. B. HARRIS

The notations employed in the equations describing the relationships applicable to various kinds of systems are not fully consistent. Since the various sets of notations have evolved as a result of work done on particular systems, and since each set of notations appears particularly suitable for its own system, the terminology is retained in this chapter in its recognized form even though some inconsistencies exist between various systems. For example, in the radar jamming problem, the symbol P_j is used to represent the power of the jamming transmitter and the symbol S_j represents the power density of the jamming signal at a point in space measured in watts per square meter. In the communication jamming case, S_t represents the transmitter carrier power of the signal, and J_t the transmitter jammer power.

13.1 Introduction

This chapter is primarily concerned with the locations of jammers and receivers and the effect of these locations on jamming parameters, including required jamming power levels. The relationship between jammer and signal powers for specific geometries can readily be derived for free space or for smooth earth. In many cases, the conditions of free space or smooth earth are not met and factors involved in the propagation of radio waves play a large part in setting limitations on usable geometric configurations between jammers and receivers.

The limitations are also affected by the particular use of the receiver being jammed, whether the receiver is being used in radars, in communications, in fuzes, or in navigation devices. Four cases, typical of the types of geometry problems which are encountered, are considered in this chapter. These cases

are airborne radar jamming, search radar jamming, communications jamming, and fuze jamming. In addition, factors, such as the radar cross section, that h directly affect the geometry problem are discussed. How propagation enters into the geometry problem is indicated, but the nature of the propagation of radio waves is not discussed since this is treated in Chapter 31.

The factors that determine jamming power at the receiver are described in the geometry of the jamming problem. Equations are readily derived in particular geometries and particular jamming situations for the transmitter jamming power required to produce a required jamming power at the receiver. The parameters entering into the equations are not so readily defined and must be carefully considered in each individual problem. As would be expected, the threshold of intelligibility or identification must be defined differently for each system and for each type of jamming signal.

It is useful to define as a standard jamming signal a random fluctuation noise which has a gaussian amplitude distribution and a rectangular power spectrum equal in width to the acceptance band of the receiver (page 5 of Reference 1). This assumption allows the utilization of the results of theoretical investigations on the signal-to-noise ratios in receivers to obtain a noise-to-signal ratio, or J/S, or equivalent parameter such as C, used in the radar jamming problem, Section 13.2), at the receiving station. J/S, which will be called the jamming threshold for the receiver, defines the ratio of jamming noise power to signal carrier power at which the standard jamming signal becomes effective against the target signal. Deviations in J/S from the standard value for other jamming signals are accounted for in some cases by the efficiency factor, M_j. This factor compares the jamming signal power required to produce the jamming threshold power level in the receiver to the power required of a signal consisting of white noise of exactly the receiver bandwidth.

13.2 The Fundamental Self-Screening Airborne Radar Jamming Equation

In developing the equation for the power required to jam a radar with a noise modulated signal at a distance R from the radar, we first consider the power density developed by the radar system at a point in space (Reference 2). It is assumed herein that the jammer is located at the target and that the distance factor which affects the attenuation of the jammer signal is therefore the same as the distance factor which affects the attenuation of the radar signal. Under these conditions, it is not necessary to consider the attenuation from the target to the radar set, and the calculations can be based upon the conditions existing in free space in the vicinity of the target. The basic procedure is to select a point P in space lying, in relation to the

target, in the direction of the radar set. At this point, the jamming power density must be large enough, in comparison with the power density of the echo reflected from the target so that when both signals reach the radar set, the jamming signal will override and obscure the echo.

Consider such a point P in space at a distance ρ from the target. The requirement for successful jamming is that, at P, the jamming signal shall be related to the echo signal by a factor C, in accordance with

$$S_j = CS_r \tag{13-1}$$

where S_j is the power density of the jamming signal at P, C is the ratio of jamming to signal power required to produce successful jamming (known as the "camouflage" factor) and S_r is the useful power density of the echo signal at point P. S_j and S_r are expressed in watts per square meter.

Considering first the quantity S_j, it is seen that this quantity is given by

$$S_j = BP_j G_j / 4\pi\rho^2 \tag{13-2}$$

where B is the bandwidth of the radar receiver in megacycles per second, P_j is the minimum power output of the jammer, in watts per megacycle, required for successful jamming, and G_j is the gain of the jamming transmitter antenna in the direction of the radar receiver. Since the total power radiated from the jammer is BP_j, the jamming power density at the surface of a sphere of radius ρ centered on the target would be $BP_j / 4\pi\rho^2$ if the jamming antenna radiated isotropically. The actual jamming power density at the surface of the sphere in the direction of the radar set will then evidently be $BP_j / 4\pi\rho^2$ multiplied by G_j as shown in Eq (13-2). Similarly, the echo power density at P in the direction of the radar receiver is found to be

$$S_r = P_e G_e / 4\pi\rho^2 \tag{13-3}$$

where P_e is the total reflected echo power radiated from the target and G_e is the "gain" of the target reflecting pattern in the direction of the radar set. In turn, it is evident that the reflected power is given by

$$P_e = S_i A_e \tag{13-4}$$

where S_i is the incident power density of the radar signal before it strikes the target and A_e is the effective intercepting area of the target in square meters. Substituting Eq (13-4) in (13-3) we have

$$S_r = \frac{S_i}{4\pi\rho^2} A_e G_e \tag{13-5}$$

and if we let

$$A_e G_e = \sigma \tag{13-6}$$

Eq (13-5) becomes

$$S_r = S_i\sigma/4\pi\rho^2 \tag{13-7}$$

Now substituting Eq (13-7) and (13-2) in (13-1) we have, as the requirement for successful jamming,

$$\frac{BP_jG_j}{4\pi\rho^2} = C\frac{S_i\sigma}{4\pi\rho^2} \tag{13-8}$$

$$BP_jG_j = CS_i\sigma$$

Since the incident radar power density is

$$S_i = \frac{P_rG_rg^2}{4\pi R^2} \tag{13-9}$$

where P_r is the peak pulse of the radar set, G_r is the gain of the radar antenna in the direction of the target, g is a factor taking into account the effect of ground echoes, and R is the distance between the radar set and the target, Eq (13-8) therefore becomes

$$BP_jG_j = C\frac{P_rG_rg^2\sigma}{4\pi R^2} \tag{13-10}$$

which, when R is converted to miles, can be expressed in the form

$$BP_j = \frac{0.0308 \times 10^{-6}P_rG_rg^2\sigma C}{R^2G_j} \tag{13-11}$$

This equation is suitable for determining the jamming power required as a function of range. Equation (13-10) can also be written as

$$R_m = 0.176 \times 10^{-3}\sqrt{\frac{P_rG_rg^2\sigma}{BP_jG_j}} C \tag{13-12}$$

a form which is suitable for determining the minimum jamming range, R_m.

As a matter of reference, the notation of Eq. (13-6) and (13-7) are now defined in the following manner:

P_j Minimum jamming noise power per megacycle bandwidth required to produce effective jamming at a range R.

B bandwidth of the radar receiver in megacycles.

P_r effective peak power of radar transmitter (actual generated power minus transmission line and other losses).

G_r power gain of radar transmitter antenna relative to an isotropic radiator, free space condition, in the direction of the target.

g ground reflection factor. (To be discussed in more detail in Section 13.4).

R distance in miles.

R_m minimum jamming range in miles.

σ cross section of target in square meters. (To be discussed in Section 13.7).

C camouflage factor, the ratio of rms noise to peak pulse power of radar for jamming to be effective. (Discussed further in Section 13.5).

G_j power gain of jamming antenna relative to an isotropic radiator, in the direction of the radar.

13.3 Maximum Detection Range

The effect of geometry on the detection range of radars in the presence of jamming can also be determined in terms of deterioration in a radar's maximum detection range and the minimum detectable signal. This determination is important when considering the capabilities of radars used in search.

If a receiver is located at the radar transmitter site, as in the case of a search radar, $\rho = R$, and Eq (13-7) becomes

$$S_r = S_i \sigma / 4\pi R^2 \qquad (13\text{-}13)$$

The echo power density S_r may be found in terms of the incident power density S_i at the target by substituting Eq (13-9) in Eq (13-7).

$$S_r = \frac{P_r G_r g^2 \sigma}{(4\pi)^2 R^4} \qquad (13\text{-}14)$$

The received power depends upon the antenna aperture which in turn is related to antenna gain by

$$A_r = G_r \lambda^2 / 4\pi \qquad (13\text{-}15)$$

where λ is the wavelength. The received power P_{rec} is the product of the antenna aperture A_r by the echo power density S_r, or

$$P_{rec} = A_r S_r = P_r G_r^2 g^2 \lambda^2 \sigma / (4\pi)^3 R^4 \tag{13-16}$$

This equation may be solved for range and for minimum received power or P_{rec} equal to P_{min}, the maximum detection range, R_{max} is

$$R_{max} = \sqrt[4]{P_r G_r^2 g^2 \lambda^2 \sigma / (4\pi)^3 P_{min}} \tag{13-17}$$

It is the minimum received power P_{min} which is affected adversely by jamming. Under jamming conditions, this power must be increased to detect the echo so that maximum detection range decreases.

Without jamming, there is a signal-to-noise ratio, $P_{min}/N_i = h$ which must be equaled or exceeded if a target is to be detected. The equivalent input noise, N_i, has a gaussian amplitude distribution. If the jamming signal is gaussian, it may be added directly to the equivalent input noise to arrive at a new noise figure. If the jamming signal is not gaussian, it must be multiplied by an efficiency factor in order to justify its addition to the equivalent input noise. If the jamming signal multiplied by the efficiency factor is designated by P_j, then a new minimum detectable signal ratio would be given by

$$\frac{P'_{min}}{N_i + P_j} = h \tag{13-18}$$

The relationship between the maximum detectable range and the received power under a non-jamming and a jamming environment can be determined from Eq (13-16) and (13-18) (page 27 of Reference 2). This relationship is

$$\frac{P_{min}}{P'_{min}} = \left(\frac{R'_{max}}{R_{max}} \right)^4 \tag{13-19}$$

Finally, the maximum detectable range R'_{max} in a jamming environment is given in terms of the maximum detectable range in a non-jamming environment by

$$R'_{max} = R'_{max} \sqrt[4]{\frac{P_{min}}{P'_{min}}}$$

$$= R_{max} \sqrt[4]{\frac{P_{min}}{P_{min} + hP_j}} \tag{13-20}$$

13.4 Ground Reflection Factor, g

The ground reflection factor, g, was introduced into the fundamental radar jamming equations in order to take into account effects resulting from the interference between the wave propagated directly to the target and the wave reflected from the ground. It is evident that interference between these two waves may increase the field at the target by a factor of 2 or reduce it to 0 (theoretically), depending on whether reinforcement or cancellation takes place. If the beam of the radar is aimed above the horizon and no ground reflection exists, then $g = 1$. If ground reflection does exist, g may theoretically be as high as 2. Since much information is now available on this subject in the literature on propagation, the actual vaules of g to be applied will not be discussed further herein, except on a qualitative basis.

13.5 The Camouflage Factor, C

The camouflage factor, as applied to pulse radars, is the ratio of rms noise power to peak signal power required to provide effective jamming (where the rms noise power takes into account the bandwidth of the receiver).

To determine the value of C to be applied in specific cases, Haeff (formerly with the Naval Research Laboratory), made studies to determine the minimum radar signal intensity which could be observed in the presence of noise using Type A presentation (Reference 3). These studies resulted in determining that the camouflage factor can be represented by the empirical expression,

$$C = \frac{4}{[1 + (1/\tau B)^2]} \left(\frac{F}{K} \right)^{1/2} \tag{13-21}$$

where

τ = pulse lengths in microseconds,

B = bandwidth of receiver in megacycles,

F = pulse repetition rate in cycles per second, and

K = an experimentally determined factor, discussed below.

Equation (13-21) shows, as might be expected, that, for a given value of B, C becomes smaller as τ becomes shorter. As the pulse lengths become shorter, the pulse amplitude in the radar receiver becomes smaller due to distortions caused by the passband being too narrow, thus making it easier to override the pulse with a jamming signal. Conversely, as B becomes larger, C also becomes larger. From this latter result, it might be expected that it would be desirable, from an anti-jamming (A-J) standpoint, to make radar bandwidths as wide as possible, to reduce pulse distortion and make C as large as possible. However, reference to Eq (13-10) shows that the minimum

jamming power required is also a function of B, and is, in fact, inversely proportional to B. Thus, although an increase in B increases the pulse amplitude in the radar receiver, it also increases the noise power accepted, and an indefinite increase in radar bandwidth is not desirable from an A-J standpoint. It can be shown, in fact, by relating Eq. (13-10) and Eq. (13-21), that an optimum A-J design results when $\tau B = 1$. The Wurzburg radars used by the Germans during the war actually achieved this optimum, while the U. S. sets, in general, had values of τB ranging from 2 to 26.

The final factor $(F/K)^{1/8}$, in Eq (13-21) represents a second order effect since variations in the pulse repetition rate cause change amounting only to approximately one decibel per octave. Both Haeff and Lawson studied the value of K, determining approximate values of $K = 1640$ and $K = 400$ re-

FIGURE 13-1   Camouflage factor vs. pulse length
τ and receiver acceptance bandwidth B

spectively. Figure 1 shows the camouflage factor a function of τB, in accordance with Eq (13-21).

Equation (13-21) does not take into account the possible existence of saturation in the receiving system. There is experimental evidence that, if the jamming signal is sufficiently strong to produce saturation, the effectiveness of the jamming is increased. This conclusion is based on actual tests of jamming signals; these tests showed that the pip could be seen more easily when the gain of the receiver was reduced, provided saturation existed initially. Thus, an important A-J measure is the incorporation of automatic gain control device functioning on noise. The question of the degree to which such expedients should be employed has been studied extensively. Data concerning the results are available in Chapter 14 references.

13.6 Gain of Jamming Antenna, G_j

The factor G_j is one which has been very extensively explored in connection with the design of jamming antennas. It may be assumed that the gains realizable under practical conditions will range from 1.64 (the gain of a half-wave antenna in free space) to hundreds or thousands depending upon the frequency and the use to which the jammer antenna is put.

13.7 The Target Cross Section, σ

The subject of target echoing areas, or cross sections, in itself has engaged the attention of a significantly large proportion of the personnel working on radar countermeasures. The echoing areas which could be expected in the case of certain types of aircraft were estimated in initial studies on this subject. These estimates were that a medium bomber had an echoing area of 50 square meters; a small fighter plane, an echoing area of 5 square meters; and a large bomber, an echoing area of 74 square meters. Since it was realized that these figures were, at best, guesses, contracts were placed with Ohio State University for studying this problem, and, in the World War II program, quite a large number of measurements of echoing areas became available as a result of the Ohio activity. As was expected, the cross section of any given aircraft was found to vary by several orders of magnitude as a function of aspect. Thus, although the original estimates of cross sections were found to be accurate for certain aspects of the aircraft, it was realized that it would be impossible to express the echoing area of any given aircraft in terms of a single figure. As a result, the observed cross sections have generally been published in the form of reflecting patterns giving the cross sections at various values of azimuth and elevation.

The selection of a value of the cross section to be used in a specific case requires taking into account the rapid variation of cross section with aspect, the extreme depth of the minima, and the extreme height of the maxima. The rapidity of the oscillations in σ for a given target depends on the wavelength, and on the separation between nulls (in aspect angle) decreasing as the wavelength decreases. Probability considerations enter into the problem at this point. It is obviously not desirable to design a jamming transmitter on the assumption that the jammer should be able to override the radar echo under conditions when the radar set lies on a maximum of the strongest lobe of the reflecting pattern. In general, this main lobe of the reflecting pattern lies at right angles to the longitudinal axis of the aircraft, and in some cases may be only 3 or 4 degrees wide, the actual width depending on the wavelength involved for the particular target. It seems obvious that it would not be economical from an engineering standpoint to design jammers to function against this lobe of the reflecting pattern, since the probability of the lobe being directed at the radar set might be as small as 1 or 2 in 100. On the other hand, it is not safe to design jammers to function against one of the pattern minima, because these pattern minima are even narrower than the maxima and because the variation in cross section between minima and maxima may be as large as 20 or 30 to 1.

Another problem entering into the calculations is the question of the effective cross section of a group of targets flying in formation. Initial studies

made by Van Vleck (Reference 2) and Norton (Reference 4), developed probability factors expressing the percent of time during which the effective cross section of the flight might be expected to exceed a value determined by the cross section of the individual airplane and the number of aircraft in the flight. These results were based on the concept that the resultant signal from a flight of airplanes may be considered to be the resultant of the individual signals added with proper attention to the relative phases (which are random), making the assumption that '.' aircraft (M in number), are spaced within half the distance occupie,   pulse in space. Since the phases are random, the extremes are represented first by the case when all echoes cancel and the effective cross section of the flight is 0, and second by the case when all echoes add and the effective cross section of the flight is M times in voltage or M^2 in power that of a single aircraft. In other words, while the average effective area of the formation may be expected to be $M\sigma$, where σ is the area of a single aircraft, it is possible for the area of the formation to vary from 0 to $M^2\sigma$. It was pointed out that the probability of any given formation cross section (Σ) being exceeded is given by

$$P = 100 \exp\left(-\frac{\Sigma}{M\sigma}\right), \tag{13-22}$$
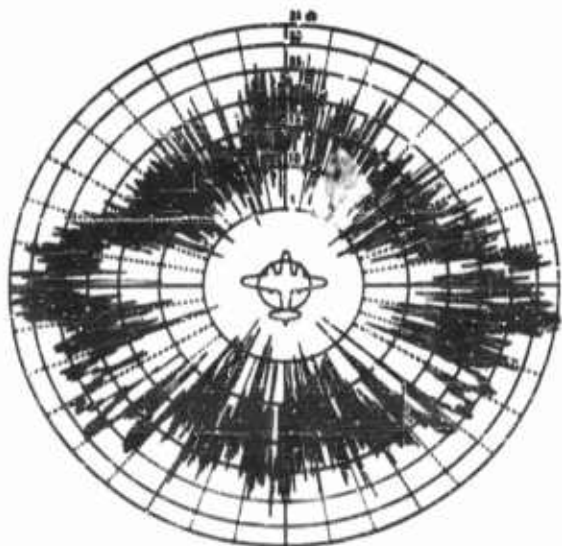
FIGURE 13-2   Echo on 10 cm from a B-26 bomber as function of azimuth

where M is the number of aircraft in the flight and σ is the echoing area of the individual aircraft. An examination of this equation shows that Σ will exceed $2\,M\sigma$ for 13.5 percent of the time, and $4\,M\sigma$ for 1.83 percent of the time. Considerable design work has been carried out on the basis that $4M\sigma$ is a reasonable value of Σ for calculations, and that when this value for σ is used protection is obtained 98 percent of the time.

When viewing aircraft flying in formation there will be a distribution in aspect around some average aspect (θ_0, ϕ_0) (Reference 5). The angle θ is measured from the zenith and the angle ϕ from the nose of the aircraft. This distribution of aspect can be expected to be essentially gaussian, but restricted for the most part to some interval, for example, from $(\theta_0 - 5^\circ)$ to $(\theta_0 + 5^\circ)$ and $(\phi_0 - 5^\circ)$ to $(\phi_0 + 5^\circ)$. Then at any given instant of time, signals will be received from the aircraft which will vary from the relative minimum values to the relative maximum values associated with the scattering pattern for a single target in the vicinity of (θ_0, ϕ_0). If the number of aircraft in the formation is reasonably large, one would expect to receive an average signal per aircraft-in-formation that could be described by a smoothed-out (i.e., averaged) pattern for the individual aircraft. This means that interest becomes centered on radar cross section patterns which represent median values over 5 degree or 10 degree intervals. Much of the experimental data obtained on aircraft is presented in this form.

The nature of the radar reflection patterns to be expected for aircraft is illustrated in Figures 13-2 through 13-7. In Figure 13-2, the extreme oscillatory nature of the radar cross section pattern as a function of aspect is seen This figure is taken from Kerr, (Page 542 of Reference 6). Figures 13-3 and 13-4 illustrate the manner in which the "average" radar cross section pattern varies with frequency. In both figures, cross sections are shown at horizontal polarization for aspects confined to the horizontal plane ($\theta = 90^\circ$). These two figures are taken from Reference 7; the first involves the F-86 and the second the B-47. In Figure 13-5, polarization effects are illustrated. The case in question is the radar cross section of the B-47 aircraft at a wave length of 4.11 meters; these experiments were conducted by Ohio State University and their experimental work on the B-47 plus other experimental and theoretical work on the B-47 is reported in Reference 8. The manner in which the cross section changes with a variation in the elevation angle can be illustrated in terms of the experimental work of Radiation, Inc. on the B-57 (Reference 9). The coordinate system employed by Radiation Inc. differs from the (θ, ϕ) system described above and thus their coordinate system is displayed in Figure 13-6. Experimental results for elevation angles ranging from 40 degrees above the horizontal to 40 degrees below the horizontal are displayed in Figures 13-7a through 13-7e. These experimental data are expressed in
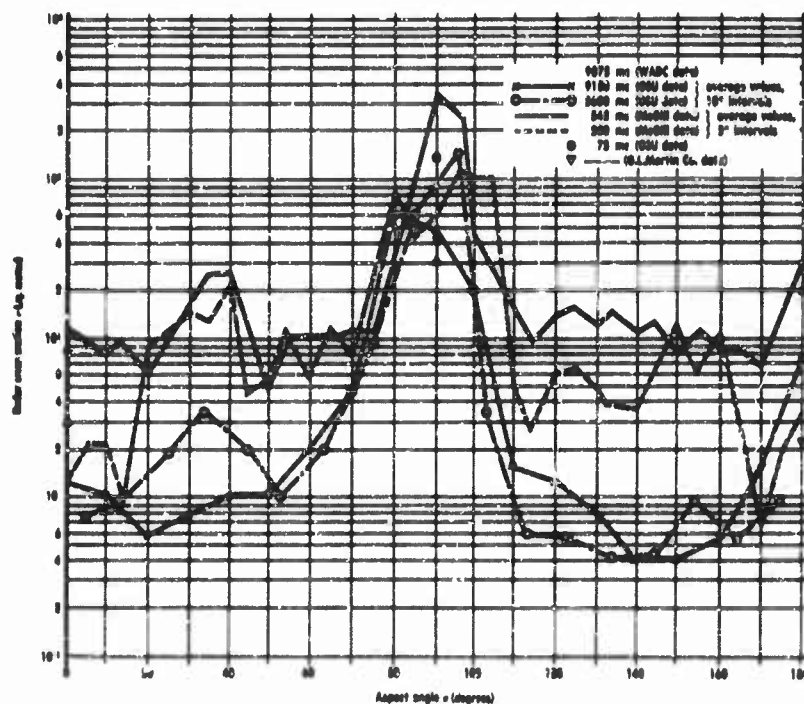
FIGURE 13-3   Cross section of the F-86 aircraft (experimental). Horizon polarization
$\theta = 90°$

terms of median and maximum values over 10 degree intervals. It will readily
be observed from these figures that there are no appreciable changes in the
patterns for this variation in the elevation angle.

Reference 7 contains, in tabulated form, a considerable amount of radar
cross section data with primary emphasis on the nose-on, broadside, and tail-
on aspects. Table 13-I is illustrative of this material. This sample of tabulated
data includes some of the available information on the B-47, the Canberra,
and the F-86. (The Canberra and the B-57 are essentially the same aircraft.)

TABLE 13-I. RADAR CROSS-SECTION DATA

Body	Equipment	Polar-iza-tion	Static or Dynamic	Frequency (mc)	c-w or Pulse	Aspect	Radar Cross Section (sq. m)
B-47 (model)	AN/APS-34	H	Static	600	Pulse	Nose-on Broadside Tail-on	25 630 *** 28
B-47 (model)	AN/APS-34	V	Static	600	Pulse	Nose-on Broadside Tail-on	50 892 *** 6.3
B-47	AN/APQ-13	V	Dynamic	9375	Pulse	Nose-on Broadside	4;150 180;15000
B-47 (model)	Hybrid T	V	Static	150	c-w	Nose-on Broadside Tail-on	5;8.6 25;250 * 1.1;1.9
B-47 (model)	Hybrid T	H	Static	150	c-w	Nose-on Broadside Tail-on	1;25 1;350 * 1;4.3
Canberra B-2	—	H	Dynamic	X-band	Pulse	Nose-on Broadside Tail-on	31.7 260 ** 14.4
Canberra B-2	—	V	Dynamic	X-band	Pulse	Nose-on Broadside Tail-on	20.2 97 ** 12.2
F-86 (model)	Hybrid T	H	Static	200	c-w	Nose-on Broadside Tail-on	1.4 55 3.8
F-86 (model)	Hybrid T	V	Static	200	c-w	Nose-on Broadside Tail-on	4.4 100 0.3
F-86 (model)	Hybrid T	H	Static	545	c-w	Nose-on Broadside Tail-on	12 300 36
F-86 (model)	Hybrid T	V	Static	545	c-w	Nose-on Broadside	9.8 500
V-formation 3 F-86's	TPS-1B	H	Dynamic	1250	Pulse	Az. 359°- 6° El. 2°-10°	1 (median)
V-formation 3 F-86's	TPS-1B	H	Dynamic	1250	Pulse	All	10
V-formation 3 F-86's	SP-1M	H	Dynamic	2810	Pulse	Az. 359°- 6° El. 2°-10°	16 (median)
V-formation 3 F-86's	SP-1M	H	Dynamic	2810	Pulse	All	50
V-formation 3 F-86's	MK-33	H	Dynamic	9380	Pulse	All	9.2

* Median and maximum values in 10° intervals.   ** Mean values.
*** 10° median values.

13-13
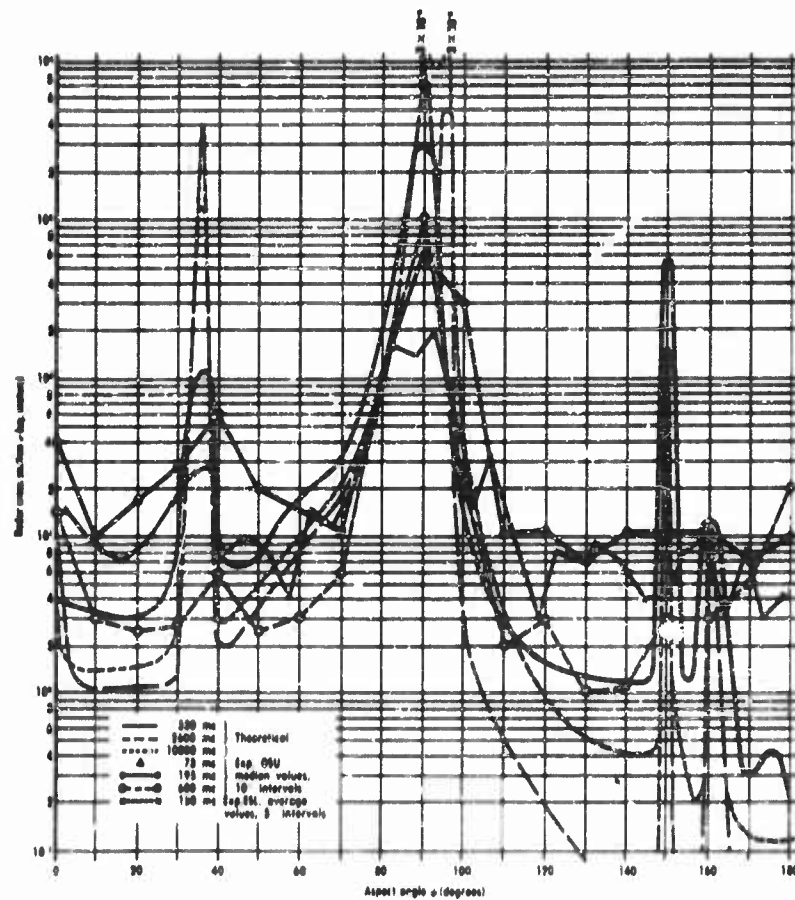
FIGURE 13-4   B-47 radar cross section at 0° elevation
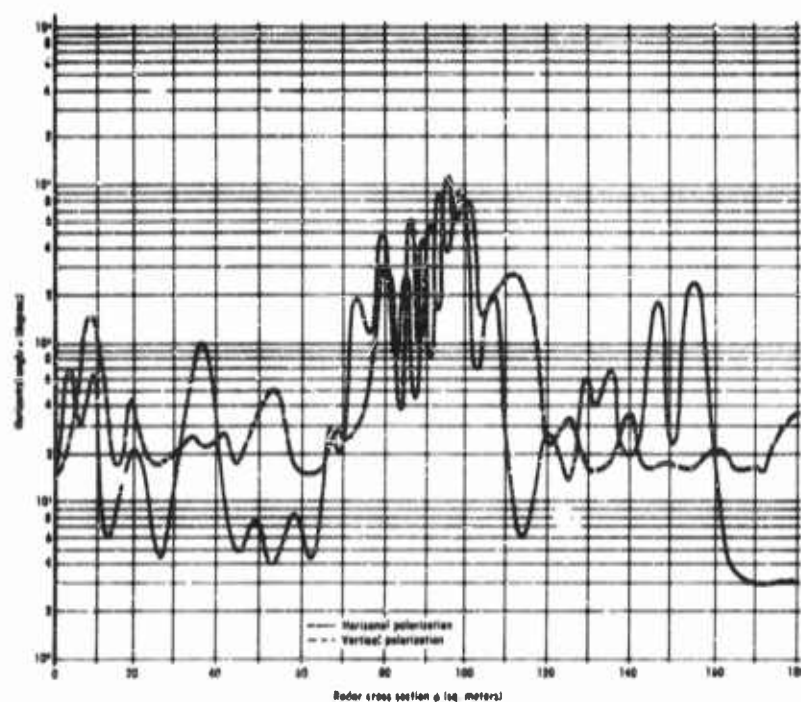
FIGURE 13-5   B-47 radar cross section, OSU experimental data. $\theta = 9C°$; $\lambda = 4.11$ meters
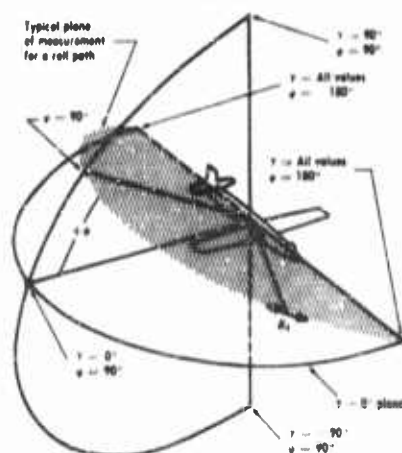
FIGURE 13-6   Measurement coordinate system
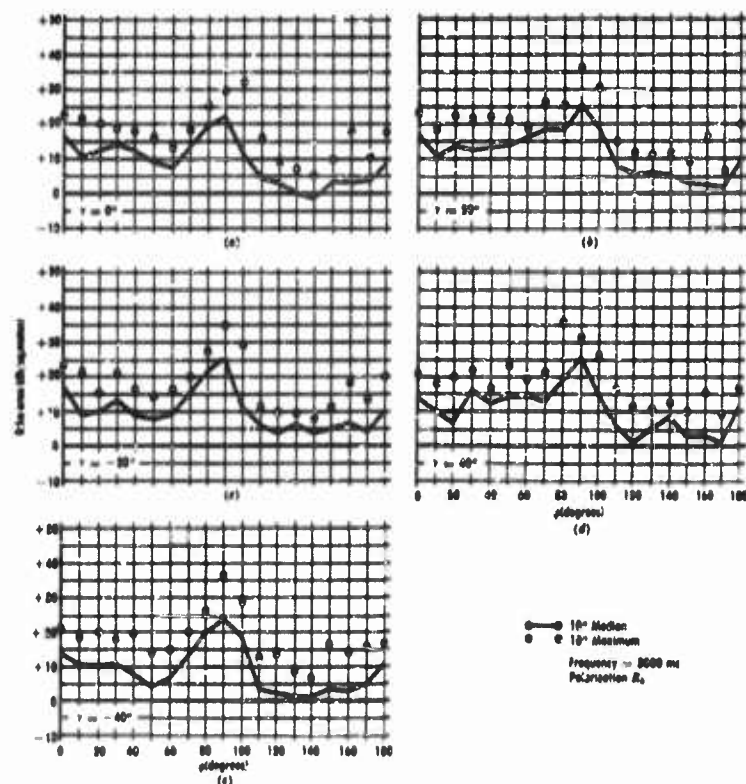
FIGURE 13-7 B-57 radar cross section experimental measurements

13.8 Communications Jamming

The following terms are defined for the case of ground-to-ground communications jamming:

S_1 = transmitted carrier power signal,

S = received carrier power of signal,

J_1 = transmitted jammer power,

J_g = received jammer power, and

J = received power of standard jamming signal.

All these quantities are measured at the antennas (Reference 1).

For the geometry in Figure 13-8, the following two equations may be written:

FIGURE 13-8 Geometry for communications jamming

$$S = S_1 \left[\frac{\lambda}{4\pi r_t} \right]_2^2 G_t G_r A_t^2 \tag{13-23}$$

and

$$J_2 = J_1 \left[\frac{\lambda}{4\pi r_j} \right]^2 G_j G_{rj} (\theta) A_j^2 \tag{13-24}$$

where

G_t = gain of transmitting antenna,

G_j = gain of jammer antenna,

G_r = gain of receiving antenna, and

$G_{rj}(\theta)$ = relative gain of receiving antenna in the direction of θ.

All gains are measured relative to an isotropic radiator.

A_t and A_j are path factors, terms which account for the discrepancies between free space values and actual values of the field intensities. These path factors, which depend upon propagation conditions, must be determined for each individual case.

If the jamming signal is restricted to the white noise standard, then Eq. (13-23) and (13-24) yield the ratio of the transmitted jammer-to-signal power is a function of J/S, where J/S, the jamming threshold, depends only on the characteristics of the receiver (including the operator) and the type of signal being received. For $J_2 = J$,

$$\frac{J_1}{S_1} = \frac{J}{S} \left[\frac{r_j}{r_t} \right]^2 \left[\frac{G_t G_r}{G_j G_{rj}(\theta)} \right] \left[\frac{A_t}{A_j} \right]^2 \tag{13-25}$$

The actual jamming signal may differ considerably from the white noise standard assumed above. In this case, the threshold J_2/S would in general also depend on the particular type of jamming signal employed as well as the characteristics of the signal and the receiver. The efficiency factor M_j can be used to account for the relative effectiveness of practical jamming signals as compared to the white noise standard so that a new J/S need not be defined for each jamming signal. For the actual jammer, Eq. (13-25) becomes for the threshold,

$$\frac{J_2}{S_1} = \frac{J}{S} \left[\frac{r_j}{r_t} \right]^2 \left[\frac{G_t G_r}{G_j G_{rj}(\theta)} \right] \left[\frac{A_t}{A_j} \right]^2 \left[\frac{1}{M_j} \right] \tag{13-26}$$

where M_j is usually less than unity but may be greater than unity in particular cases and J/S is defined on the basis of white noise and is independent of the type of jamming signal.

13.9 Radio Proximity Fuze Jamming

The geometrical picture of the fuze jamming problem is more complex than those of the communications and radar problems. In the fuze jamming problem, propagation effects, however, are confined to those which affect transmission characteristics at short ranges.

The doppler proximity fuze functions upon approaching a target when the reflected signal reaches a specified fraction of the radiated signal. The reflected signal appears to the fuze to be at a slightly different frequency from the radiated signal due to the motion of the fuze toward the target. This doppler difference in frequency (a few hundred cycles) appears as a relatively low-frequency amplitude modulation. The doppler frequency variation is amplified and, upon reaching a certain prescribed level, fires the fuze.

Jamming a fuze is accomplished by presenting it with a jamming signal stronger than the actual reflected signal before the missile reaches its target. Fuzes can be made to discriminate against certain types of jamming signals; so that the effectiveness of a jamming signal must therefore be described for a particular fuze. The calculation of the total power requirements for any given tactical situation depends primarily upon the number of jammers required to cover the expected fuze frequency band and the power required for each jammer to perform its function. This latter quantity involves the characteristics of the fuze and jammer, the trajectory of the missile, and the target to be protected.

Although several jammers may be required to provide protection over the entire frequency range, it may be assumed that sufficient power can be given to one jammer to cover one interval of that range. It will be this power per interval that will be referred to in the following discussion. This power is determine the geometry of the target to be protected, the physical properties of the attacking weapon, and the countermeasures device.

The effect of geometry on the operation of fuzes has been considered for many cases including use of fuzes against airborne targets (Reference 10) and use of fuzes against ground targets. The following discussion is concerned with ground targets. The term protection refers to the prefunctioning of fuzed missiles on or above the horizontal plane (the ceiling) at a specified vertical distance (ceiling height) above ground. Level ground is assumed. At any point on the ceiling the field strength measured by the fuze, called the effective field strength, must be sufficient to jam the fuze. The locus of points at which the effective field strength is equal to the jamming field strength is called the burst surface.

For protection of a ground area, the burst surface must intersect and extend above the ceiling, thus intercepting a certain ceiling area, in which the requirements of ceiling height and sufficient field strength to jam the fuze are both met (Figure 13-9). The intercepted ceiling area and the burst surface above the ceiling area projected along the missile trajectory onto the ground determine the protected area and the shadow area respectively. Both areas provide protection. When the missile comes straight down, the shadow area merges into the protected area. (Reference 11).
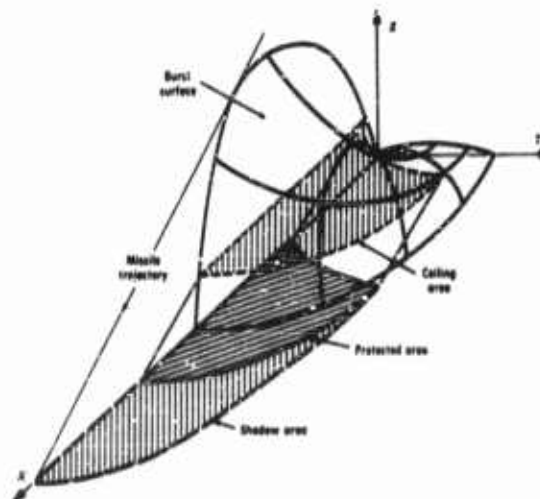
FIGURE 13-9 Ceiling area, ground area protected by it,
and additional shadow area

Jamming power as a function of geometrical conditions depends upon the radiation patterns and polarization of both the jammer and the fuze antennas; it must be determined separately for each basic jammer-fuze antenna combination. In general, the method is not analytic but requires graphical or numerical computation. (Reference 12). The following antenna combinations have been considered in a study by the Electronic Defense Group of the Engineering Research Institute of The University of Michigan (Reference 11):

a) Vertical jammer antenna, longitudinal fuze excitation,

b) Horizontal jammer antenna, longitudinal fuze excitation,

c) Vertical jammer antenna, transverse fuze excitation,

d) Horizontal jammer antenna, transverse fuze excitation,

e) Discone antenna with a corner reflector, longitudinal fuze excitation,

f) Discone antenna with a corner reflector, transverse fuze excitation.
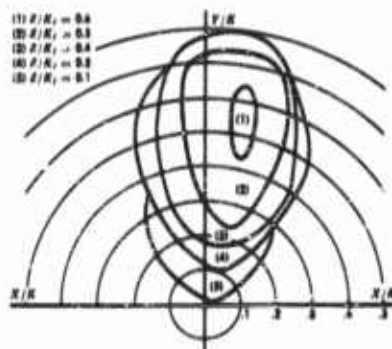
FIGURE 13-10 Protected area—AS-542 discone antenna with corner reflector, longitudinal fuze excitation

A typical example of the results of calculations is shown in Figure 13-10. The missile is acting as the fuze antenna and is longitudinally excited. The AS-542 discone antenna with corner reflector is assumed for the jammer. In this example, the normalized protected area is shown for a missile arriving from the negative X direction at an angle of 30° with the X-Y plane. K_I is a normalizing parameter and Z is the height of the ceiling area (Figure 13-9) above the ground plane. The protected area is the projection of the ceiling area on the ground. Shadow area, also covered, is the projection on the ground of the burst surface above the ceiling. Note that in Figure 13-10 the covered areas are shifted to the right of the Y axis since the fuzed missile is coming from the left. Also note that the area occupied by the jammer is not covered.

Z is generally chosen to be 1000 feet. If Z/K_I is chosen to be 0.1, the distance between the circles in the X-Y plane is 1000 feet also. If Z/K_I is chosen as 0.05, the distance between the circles is 2000 feet. The 0.05 and 0.1 values of Z/K_I lie within the range of values to be expected from a swept jammer like the AN/MRT-4 (the AS-542 discone antenna is used with this jammer) which is used against howitzer shells.

13.10 Other Aspects of the Geometry of the Jamming Problem

The airborne jamming problem discussed in Section 13.2 exemplified the method of attack required in most problems involving the calculation of jamming power or minimum jamming range. It is, however, a special case in several respects. In the first place, the assumption is made that the jammer is carried in the aircraft being screened. Second, the effect of the ground is considered only in a qualitative manner. Third, the results are dependent on the assumption that the target, being small in extent, occupies only a fraction of a lobe of the radar antenna beam.

Kuhn and Sutro made a more exhaustive and precise study of the jamming problem in Report 411-93, "Theory of Ship Echoes as Applied on Naval RCM Operations" (Reference 13). This report was intended to apply only to the screening of ships, and much of the rather complete analyses contained in it was necessitated by the fact that the assumptions of the airborne case cannot be used; nevertheless, much of the information in Reference 13 is applicable to the jamming problem in general. In particular, Kuhn and Sutro have developed relationships which permit the target to intercept more than one lobe of the radar antenna. These relationships take precise account of the interference between the direct ray and the ground reflected ray, and also cover the case when the jammer is not carried in the target.

Another type of geometry problem which was not discussed in this chapter is the determination of the area or volume of protection afforded by a multiple number of jammers. As examples of this type of problem, the use of jammers against artillery shells (Reference 14) or the use of jammers by aircraft flying in formation against air-to-air missiles may be considered (Reference 10). In these cases, the patterns of the jammers' antennas, jammer powers, the spacing between jammers, the direction of arrival of the missile, and terrain features (if applicable) must be taken into account.

REFERENCES

1. Welch, H. W., et al., "Study of the Feasibility of Ground to Ground Jamming," Interim Report, Task Order No. EDG-5. The University of Michigan, Engineering Research Institute, Electronic Defense Group. February 1952. (SECRET)

2. Terman, F. E., and White, W. D., "Notes on Power Required for Noise Jamming," Radio Research Laboratory Report No. 411-44, 18 June 1943.

3. Haeff, A. E., "Minimum Detectable Radar Signal," Naval Research Laboratory, Report No RA 3A 208A, 23 February 1934.

4. Landon, V. D. and Norton, K. A., The Distribution of Amplitude with Time in Fluctuation Noise, *Proc. IRE*, Vol. 30, pp. 425-429, September 1942.

5. Hamermesh, M., "Analysis and Application of Measurements of Radar Cross-Section of Airplane Models," Radio Research Laboratory Report 411-157, 12 February 1945.

6. Kerr, D. E., "Propagation of Short Radio Waves," McGraw-Hill Publishing Company.

7. Siegel, K. M., Burdick, W. E., Crispin, J. W., and Chapman, S., "Studies in Radar Cross Sections -- XX: Radar Cross Sections of Aircraft and Missiles," The University of Michigan, Radiation Laboratory Report, published in ONR Symposium Report ACR-10, 1956. (SECRET)

8. Schensted, C. E., Crispin, J. W., and Siegel, K. M., "Studies in Radar Cross Sections -- XV -- Radar Cross Sections of B-47 and B-52 Aircraft," The University of Michigan, Radiation Laboratory Report No. 2260-1-T, August 1954. (CONFIDENTIAL)

14

Effectiveness of Jamming ⌐ ⌐als

R. H. BENNIGHOF,   H. W. FARRIS,   L. K. LAUDERDALE,
R. H. RICHARD,   T. A. WILD

14.1 General Considerations Regarding Jamming Effectiveness

14.1.1 The Concept of Jamming Effectiveness

In order to give a useful meaning to the concept of jamming effectiveness, it is necessary to specify the environment and conditions under which the jamming effort is being pursued. Although it is conceivable that a field situation may exist wherein one jammer is directed against one particular electronic equipment, in general the ECM picture is not so simple. In most situations a vast array of ECM equipment is directed against a complex electronic system which includes, for example, radio communication channels, search radars, and tracking radars.

Obviously it would be impossible to predict precisely the over-all effectiveness of ECM in a situation of this nature. It would require thorough knowledge of many complex factors such as personnel training, morale, fatigue, weather conditions, etc., in addition to the purely electronic factors. It seems necessary therefore to evaluate jamming effectiveness in terms of the relative success of a jamming technique against a specific class of electronic equipments. This evaluation is sufficiently general to give guidance in the selection and apportionment of jamming equipments, the choice of which is also based on knowledge of the electronic system that is to be jammed and the prevailing tactical situation.

14.1.2 Obscuring versus Deceptive Signals

In the usual sense of evaluation of jamming effectiveness, we refer to a measure which does not involve saturation of the enemy's channel by brute-force techniques. That is, while his signal may be overwhelmed, his receiver is not. It is usually necessary to be more sophisticated, partly because he is likely to have a distance advantage to begin with and may, in addition, be using a transmission method which gives him a peak-power advantage. For example, the first case usually obtains in the communication situation where the jammer cannot hope for more than equal footing; i.e., both the jammer and the target transmitter encounter the same path attenuation when radiating toward the receiver. Again, in the radar case, the target faces the inverse fourth-power law of attenuation while the jammer, with its one-way transmission, may enjoy the inverse square law. However, as an illustration of the peak-power advantage, the latter example must be examined again, for the radar may use a high peak power with a low duty cycle, where the jammer may very well have to operate on a continuous basis.

The above considerations lead us to examine means by which we may best make use of our available jamming power, assuming that brute-force type powers will not be available to us. There are two distinct ways in which the enemy can be denied effective use of his electronic equipment. The first method, called obscuration, consists of transmitting a jamming signal of such power and composition that the enemy's electronic data will be completely submerged in the interference. In the other method, called deception, the jamming consists of false signals that have similar characteristics to the enemy's electronic transmissions. These deception signals either cause the enemy to select the wrong signal, or saturate his facilities to the extent that no sound choice can be made.

Obscuration techniques, which deny all electronic information, except perhaps the fact that a jammer is operating, are clearly more desirable from the standpoint of completely nullifying the enemy's use of electronic equipment. However, assuming that the enemy's electronic systems make intelligent use of frequency and space diversity, it will not, because of practical limitations introduced by the power, weight, and quality of jamming equipment required, be possible to rely on obscuration techniques alone. Deception techniques can be expected to play an important part in most jamming efforts.

14.1.3 Laboratory Evaluation versus Field Evaluation

The evolution of an idea for a jamming equipment is usually followed by a paper study to determine feasibility and optimal design. After the equipment has been constructed there arises the question of how to evaluate its

ability to perform the intended function. Since the device is ultimately to operate against electronic equipments under field conditions, it would appear, at first perusal, that the only realistic way in which the performance of jamming equipment can be determined is by field evaluation. In general however, because of the man, practical difficulties encountered in field evaluations, it is necessary to combine both laboratory and field evaluations to insure adequate measurement of the characteristics and performance of jamming equipment.

In field testing it frequently becomes extremely difficult to separate the effects of many related factors on jamming effectiveness. The compilation of sufficient data to gain a thorough understanding of all of the factors related to equipment performance often becomes impractical in view of the expenditure of time and money required. By the use of simulation techniques, it is normally possible to study jamming equipment performances in the laboratory with a precision that is virtually impossible in the field. Information obtained in laboratory studies can then be used to design a realistic and efficient field evaluation program.

14.1.4 Relationship Between Jamming Effectiveness and Susceptibility to Jamming

A natural consequence of the development of a catalog of jamming signals and their relative effectiveness measures is a determination of the susceptibility of certain equipments to jamming. It might at first appear that the logical organization to determine jamming effectiveness is also the one to establish the measure of susceptibility. While this may be an economical move, it is one which may be a handicap to the entire assignment.

To determine the effectiveness of various jamming signals, the engineer would like to have available to him a "universal" receiver. It is representative of all receivers and does not embody the peculiarities of any one. It has large dynamic range, is linear where it should be so, does not suppress weak signals or noise in any other than theoretical fashion, receives all types of signals and demodulates them in an optimum manner and permits variation of its parameters, such as gain, bandwidth, time constant, and the like with complete flexibility. With the receiver in hand, he is free to compile a catalog of the relative effectiveness of a wide variety of jamming signals, knowing that their ordering will be independent of any receiver peculiarities. Then, with such a catalog of signals and the appropriate modulations available, he is prepared to determine the susceptibility of less-than-ideal receivers to the various jamming transmissions which can be impressed upon their antenna terminals.

In the absence of the universal receiver, or ideal receiver, the specification

of both jamming effectiveness and susceptibility of receivers to jamming be-
comes an "if-then" proposition—"*If* our receiver and target signal para-
meters are as given, then the effectiveness of FM-by-noise jamming signal A
is three decibels better than AM-by-noise jamming signal A." Or, "If an
FM-by-noise jamming signal A of *x*-kc deviation at *y*-kc modulation band-
width is used, then our receiver, when tuned on frequency to signal B shows
a higher degree of susceptibility than when subjected to AM-by-noise jam-
ming signal A." Such expressions and qualifications lead one to present his
data in the form of curves, rather than strive for an unrealistic single num-
ber. The situation is not unlike that of a manufacturer-customer relationship
in an equipment transaction. As can readily be appreciated, a manufacturer
does not present a number which evaluates the equipment in terms of a par-
ticular use the customer has for the product. Rather, he presents a set of
data which permits the customer to calculate the value of the product for
his own use. Such should be the case in presentations of studies of both
jamming effectiveness and susceptibility to jamming, but the clear specifica-
tion of the conditions is mandatory because of the very complete inter-
dependence of the one study on the other.

14.2 The Search Radar Jamming Problem

14.2.1 Introduction and General Comments

Since the end of World War II there have been many technological ad
vances that have contributed to the improvement of search radar perform-
ance. Microwave power sources capable of increasing the radar radiated
power by orders of magnitude have been developed. Also, basic radar cir-
cuitry and components have been vastly improved in both performance and
reliability. New circuits have been added to enhance the flexibility of radars
operating in a variety of interference and jamming environments and to
insure optimum processing of radar data. However, the most significant
single new search radar innovation, with regard to decreasing vulnerability to
electronic jamming, has been the acquisition of a rapid tuning capability.

Formerly the problem of jamming a search radar was relatively simple. It
was only necessary to tune the jammer to the radar frequency and transmit
relatively little power to achieve successful jamming. The radar was con-
strained by the inherent characteristics of its microwave power source to
remain at a fixed frequency; consequently all of the jamming power could be
concentrated in the radar receiver bandpass, thus producing a large jam-
ming-to-signal power ratio. Under these conditions of "brute force" jamming
the type of jamming waveform employed was of little consequence and was
usually chosen according to convenience of jamming equipment design.

Now, the frequency agility of modern search radars requires that the jamming power be distributed over the entire frequency-operating range of the radar in order to jam successfully. This means that the actual jamming power in the radar passband has been reduced by a factor of several hundred. The jamming to signal ratio in the radar receiver has also been reduced accordingly, with increased radiated radar power resulting in an even further reduction.

From these considerations it is clear that the choice of an optimum jamming waveform is central to the problem of maximizing jamming effectiveness. In general it can be stated that the jamming waveform must satisfy three essential conditions in order to produce successful jamming: (1) sufficient power, (2) uniform frequency coverage, and (3) a time structure of sufficient complexity to obscure target echoes or to produce numerous false targets that can not be distinguished from the true target echo.

Theoretical Evaluation of Jamming Effectiveness. It was pointed out in 14.1.1 that a comprehensive analysis of the jamming effectiveness problem is virtually impossible because of the many interrelated influencing factors. This does not mean, however, that theoretical analyses can not prove extremely useful in providing an understanding of how various parameters influence ECM effectiveness. It is inconceivable that an experimental investigation could be conducted, even in the laboratory, in a finite amount of time that would produce complete data concerning the effectiveness of all possible modulations against a search radar system. Therefore it is imperative that mathematical models of the jamming problem be constructed in order both to guide experimental studies along fruitful paths and to interpret the significance of experimental data.

The problem of detecting radar signals in gaussian (or receiver) noise has been treated by Lawson and Uhlenbeck in Reference 1, Marcum in Reference 2, and others. Hok (Reference 3) has applied the principles of information theory to the jamming problem to show that maximum equivocation of the radar receiver's output is obtained using bandlimited gaussian noise. In order to bring in the decision process explicitly, which is central to the problem of radar detection, the methods of statistical decision theory (Reference 4) must be applied. Here a cost criterion is chosen, and the optimum jamming signal is defined as the one that maximizes the average cost of the decision that the observer makes when subject to jamming (Reference 5).

It has been stated previously that, in addition to the power and frequency spectrum of the jamming signal, the time structure of the jamming waveform is directly related to jamming effectiveness. The former two properties are adequately described using second order statistics, as, for example, will

be seen in the calculations in 14.2.4. However, since spectral analysis does not preserve the phase information which is closely related to the waveform time structure, a further mathematical description of the jamming waveform is needed. Middleton has suggested the use of complexity measures proportional to the natural logarithm of the probability densities (Reference 5).

When a human operator is employed as the decision making apparatus, as opposed to an automatic search radar system such as SAGE, the rather indeterminate nature of the eye-brain process in the observer makes a complete theoretical analysis impossible. It is necessary to supplement the mathematical description of the man-machine system with data from carefully designed experiments. Accordingly it will be useful to review the essential features of some psychological studies in search radar visibility.

Psychological Studies of Search Radar Visibility. A rather complete survey of numerous psychological studies related to the search radar visibility problem has been compiled by Baker and Thornton, and Williams (References 6 and 7) has reviewed research in radar visibility using both deflection modulated and intensity modulated displays. An analysis of the major factors affecting visibility on intensity modulated radar displays was given by Morgan (Reference 8). This section will be concerned primarily with considerations advanced in the latter reference.

There are many variables that affect the appearance of the radar display, such as pulsewidth, pulse repetition rate, jamming spectrum, antenna beam-width, antenna rotation rate, cathode ray tube (CRT) bias, etc. However the observer is not directly concerned with these variables. His objective is to detect a target in the resulting display; consequently the visual factors, which include background brightness and composition, incremental target brightness, and target size and duration, will directly determine the observer's detectability threshold.

Assuming that the display background is relatively uniform (this would be the case for obscuration jamming), the task of the observer becomes one of brightness discrimination. The target produces an incremental brightness $\triangle I$, which adds to the background intensity I. The threshold is conveniently expressed in terms of the ratio of $\triangle I$ to I for a particular I. Because of the large range of this ratio, the logarithm of $(\triangle I/I)$ is used. A set of curves depicting typical brightness discrimination threshold is given in Figure 14-1.

The background brightness of a cathode ray tube is a direct function of the tube bias. Figure 14-2 illustrates this relationship at the time of peak excitation and at a time one-sixth second later. When a signal is applied to the grid of the CRT, it produces a change in the bias voltage for the duration of the signal pulse, and thereby produces an incremental brightness $\triangle I$.
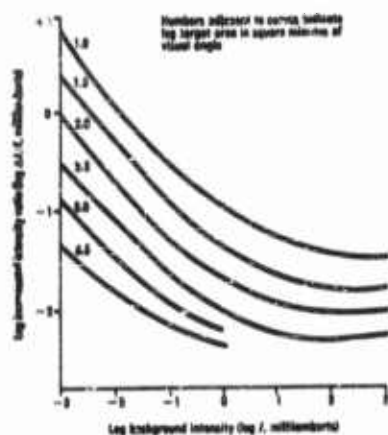
FIGURE 14-1   Brightness discrimination
thresholds
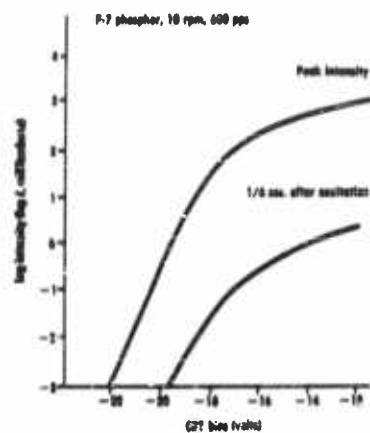
FIGURE 14-2   Light output versus cathode
ray tube bias
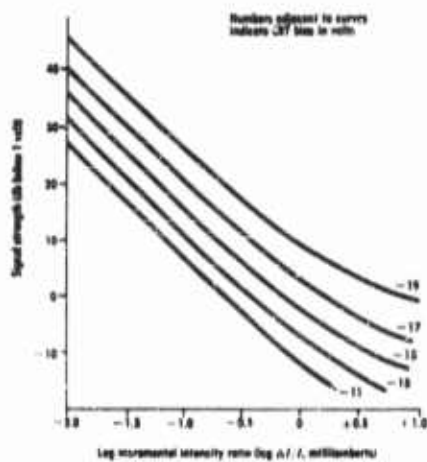
FIGURE 14-3   Signal strength versus incre-
mental brightness
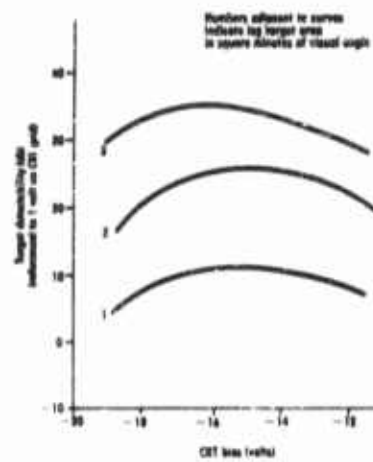
FIGURE 14-4   Target detectability versus
cathode ray tube bias

With the aid of Figure 14-2 a curve can be constructed, Figure 14-3, which shows the relationship between $\triangle I/I$ and the signal voltage on the grid of the CRT. The curve in Figure 14-2 for a time one-sixth second after peak excitation was used in constructing Figure 14-3, because the eye reaction time for maximum sensitivity is on the order of one-sixth second.

Perhaps the most important variable affecting target detectability, over which the operator has control, is the CRT bias. As may be seen in Figure 14-1, the human eye detects increments in brightness more readily as the background intensity is increased. On the other hand, it may be seen from Figure 14-2 that as the bias becomes less negative (corresponding to increasing I) the slope of the curve, which is a measure of incremental brightness $\triangle I$ per unit target signal voltage, becomes smaller. Thus, detectability at low intensities is impaired by the characteristics of the eye, and at high intensities by the characteristics of the cathode ray tube.

Assuming a given target size (the size will depend on antenna beamwidth, target pulsewidth, target range, and the distance of the eye from the scope face) a plot of detectability versus CRT bias can be constructed. This is done for targets of three different sizes on Figure 14-4. It can be seen that an optimum CRT bias, which depends somewhat on target size, exists. The calculated curves of Figure 14-4 compare quite well with the experimental data given in Reference 7. Noise jamming which produces uniform background brightness will cause the actual optimum bias to be shifted in proportion to the rms value of the noise.

Other important radar parameters that affect the brightness of the cathode ray screen are the antenna rotation rate and the pulse frequency. In order to produce a uniform background these two parameters must be compatible, that is, the pulse repetition frequency must be sufficiently high for a given antenna rotation rate so as not to produce spokes on the screen. The background brightness will vary inversely with the antenna rotation rates, and will be directly proportional to the pulse repetition frequency. By referring to Figure 14-2, which corresponds to a repetition frequency of 600 pps and a rotation rate of 10 rpm, one can calculate the effects on I of changes in the repetition frequency and rotation rate. Having determined the brightness in this manner, the required value of log $(\triangle I/I)$ may be found.

Target pulsewidth and antenna beamwidth enter the detectability problem through their effects on the target size. By using methods similar to those of the preceding paragraph, the effect of these two parameters may be calculated.

From this brief description of psychological studies of radar visibility, it can be seen that considerable progress has been made in gaining an understanding of factors governing a radar operator's performance. Under certain

conditions detection thresholds can be predicted with reasonable accuracy. However much work remains to be done in studying the effect of noise which produces a nonuniform background. In this connection, the notion of complexity measures, referred to earlier, may prove beneficial in specifying jamming noise in terms of the display it produces.

14.2.2 Types of Obscuration Jamming

Obscuration jamming signals can be conveniently classified according to the ratio of the jamming signal bandwidth to the acceptance bandwidth of the victim radar. If the ratio is large the signal is called barrage jamming; if the ratio is small the signal is called spot jamming. Spot jamming sources will not be discussed here, since they would not be effective (even when controlled by rapid automatic search-and-lock equipment) against modern search radars having pulse-to-pulse frequency-shift capabilities. As mentioned before, it is not practical to cover the spectrum with barrage jamming. However, broadband barrage jammers can be provided with electronic tuning, and one or more jammers can be deployed in frequency so as to best meet the existing search radar threat.

The class of barrage jamming signals can be further subdivided according to whether the modulation is periodic or random. Periodic modulations are unreliable because of the likelihood of synchronism or near synchronism, of unknown phase, between the jamming and echo signals in the victim radar receiver. The performance of random jamming is thus more accurately predictable than that of periodic jamming. Furthermore, a sufficient increase in jamming power will nearly always suffice to obscure the target echoes if random jamming is used, but not if periodic jamming is used.

Random Barrage Jamming Signals. The major random barrage jamming signals of established general effectiveness against search radars are direct noise amplified (DINA), FM-by-noise, and AM-by-noise. In addition, random pulse amplitude modulation can be applied to any of these signals. The salient features of each of these signals are discussed below.

Direct noise amplified is simply bandlimited gaussian noise. Since we are discussing barrage sources, it is implied that the spectral density is nearly constant over the victim receiver's passband; for practical purposes, then, DINA may be considered to be "white". Since it is the same as thermal agitation noise except for power level, its jamming effect is a large increase in receiver noise figure. DINA is the classical form of jamming, and has the most general utility of the known jamming signals. Although it may not be the best jammer in any specific situation, it is a good jammer in nearly all situations. In practice DINA is generated by amplifying low level noise

that has been filtered to obtain the desired jamming frequency spectrum. The output stage consists of a power amplifier of the traveling wave or distributed amplifier type.

Frequency modulation by noise is best discussed after categorizing into FM-by-WB (wideband) noise and FM-by-LF (low-frequency) noise. The jamming mechanism is quite different for the two cases, as will be shown in the analyses of 14.2.4.

Frequency modulation by WB noise attempts to produce the same result as does DINA, using a rapidly tunable power oscillator, such as the backward wave oscillator (BWO). A critical comparison of the relative merits of power amplifiers versus power oscillators as barrage jamming sources will not be attempted here. Selection of the best jamming source will obviously depend on such factors as the relative size, weight, cost, and reliability of the power tubes that are available in the frequency range of interest. It is of interest, however, to investigate the mechanism by which FM by noise techniques can be used to produce jamming that is essentially indistinguishable from DINA at the output of a given radar receiver, and to determine the requirements that must be placed on the FM modulation parameters. Each time the frequency modulated carrier sweeps across the victim's passband, the victim receiver's filter circuits are set to "ringing" by the impulsive character of the input. If the modulation is random, then the receiver input is a random time series of short pulses. If, further, the average frequency of these pulses is much greater than the victim bandwidth, then the conditions for the Central Limit Theorem are approximated, and the output of the receiver filter (usually i-f) is very nearly gaussian in its first order statistics. Thus, one expects the i-f output for FM-by-WB noise to be the same as for DINA.

Frequency modulation by LF noise uses the same microwave sources for its generation, but restricts the modulating noise bandwidth to be much less than the victim bandwidth. Thus, the ringing caused by one receiver crossing is usually nearly over before another crossing occurs. The i-f output wave is therefore a random time series of distinct pulses whose duration is approximately the reciprocal bandwidth. This jamming, when directed against search radars, exhibits two principal advantages and one principal disadvantage. It has increased effectiveness because the ordinary radar second detector produces more video power for a given i-f output (or receiver input) power with FM-by-LF noise than with FM-by-WB noise or DINA. Thus, this source is more efficient in producing video jamming than are the others. Also, a given video power is more effective in jamming small target displays on a PPI if FM-by-LF noise is used. This may be associated with a confusion effect caused by the resemblance of many of the bright spots to small target

echoes. The principal disadvantage of FM-by-LF noise is that it is relatively easy to counter, since the jamming is discontinuous even at the receiver output, and many or most of the target echo pulses are free of jamming if observed in real time.

Amplitude modulation by noise enjoyed considerable early usage, and approximates DINA in effect. However, it is difficult to produce a broad barrage, since the frequency coverage from a single AM-by-noise source is limited to twice the bandwidth of the modulating noise.

In randomly pulsed barrage (RPB) jamming, a background level of jamming is maintained, and the level is increased by the pulse modulation. The duration, amplitude, and spacing of the modulating pulses should be varied randomly to prevent easy countering by the victim. The average jamming pulse duration should match that of the expected victim radars, and the average jamming prf must be much greater than the radar prf. The barrage jamming signal before pulse modulation can be either DINA or FM-by-WB noise.

RPB jamming achieves high effectiveness due to the intermittent character of the jamming in the victim receiver output, in the same manner as FM-by-LF noise. In addition, it is difficult to counter since the background jamming is continuous. This jamming technique is best applied by pulse modulating in a tube having a higher peak than average power rating.

14.2.3 Experimental Methods

As discussed in 14.1.3, the complexity, expense, and inaccuracy of field testing often necessitates laboratory simulation testing of jamming effectiveness. This is particularly true when a human operator is the decision making link, as in the case of search radar using PPI display. Accordingly, it will be valuable to discuss experimental methods of evaluating the effectiveness of barrage jamming signals against PPI search radars.

Practical Criteria for Jamming Effectiveness. The first problem in designing a test is the establishment of a practical criterion to use as a measure of jamming effectiveness. As mentioned before, the enormously complicated electronic warfare situation in a hot war environment forces one to the artifice of studying conflicts between a single target-borne jammer and a single radar. For this case, a reasonable effectiveness criterion from the jammer designer's point of view is the jamming power required to reduce detection probability to a given value. In fact, the entire function relating required power and detection probability is even more useful. Of course this power is also a function of target characteristics, jamming antenna gain, range, and radar parameters. The first three of these vari-

ables can be eliminated from consideration in the laboratory by substituting radar input signal-to-jammer (S/J) ratio for jamming power, as a criterion. We have then for a criterion for jamming ratios of signal effectiveness against a particular radar, the ratios of a signal power to jamming power (per megacycle) at the receiving antenna, corresponding to various values of detection probability. To insure reproducibility and comparability of test results, the test conditions, including selection and conditioning of observers, must also be specified. These conditions will be discussed later.

Since detection thresholds are involved in the effectiveness criterion selected, then in effectiveness testing the methods used by psychophysicists in determining sensory perception thresholds are applicable (Reference 9). The experiment should proceed from easy to difficult in discrete steps, and the forced choice method, where the observer is not permitted to reply "I don't know", has been found necessary for reproducibility of results. In addition, care must be taken to select test conditions which are reproducible, and to be consistent in these conditions. Descriptions of testing details form an important part of the final data.

As will be seen, further artifices are dictated by experimental convenience. These, added to the complexity and variability of even the simplest field situations, make it impractical to attempt accurate predictions of jamming performances on an absolute basis. What are more within reach are estimates of jamming performance relative to some standard signal. Its versatility and the large volume of data on its performance lead to the selection of DINA for that standard source. So we measure, principally, *relative* jamming effectiveness between different sources. This is expressed as the ratio of jamming power levels for equal detection probabilities.

Test Conditions and Procedures. In the development of a testing procedure, it remains to adapt the selected effectiveness criterion to the realities of the laboratory. The duration of the jamming tests must be minimized for two reasons. First, a progressive jamming test is fatiguing to a conscientious observer. Second, the cost of data in terms of manpower and equipment is high (though still much lower than the cost of equivalent field test data). The designer of a jamming effectiveness experiment is called upon to make a very careful compromise between cost on the one hand and accuracy and realism of the data on the other hand.

In a conflict between a single target-borne jammer and a single search radar, the jammer's mission is to obscure range information, since azimuth information is provided by the jamming source. In a forced-choice simulation experiment, then, it would be natural to fix the target azimuth, and to require the observer to state at which of several possible range positions the

target is located. However, on a PPI the detection probability varies with range, because of the variation of spot size effects and of subtended visual angle. The added experimental complication of this effect is avoided by fixing target range and requiring the observer to state at which of several possible azimuth positions the target is located.

A further artifice of interest is the elimination of antenna scanning modulation of the jamming noise; i.e., the noise is presented uniformly over the entire PPI screen. The purpose is a drastic reduction in observer fatigue. The justification is the fact that relative effectiveness is being measured. By the same argument, complex antenna pattern modulation need not be impressed on the simulated target echo; the use of rectangular gating waveforms simplifies the equipment problem. Note that whenever an experimenter uses this justification, he is assuming that the artifice introduced has equal effects, quantitatively, on the jamming sources being compared.

We have now outlined criteria and the essential features of a method for laboratory testing of relative jamming effectiveness against a PPI type search radar. Some artificial experimental conveniences have also been described. An extensive program at the Johns Hopkins University Radiation Laboratory has been conducted with the above as a basis (References 10, 11, and 12). The other salient features of this program will be described below, to complete an example of experimental methods.

The tactical model employed is an aircraft at a fixed range of 25 (nauti-
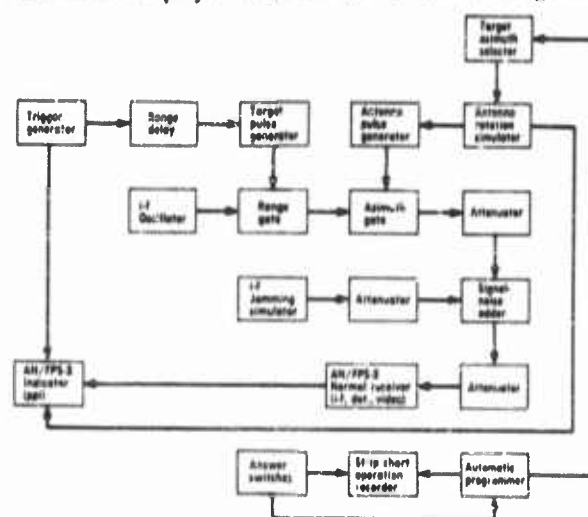
FIGURE 14-5   Jamming effectiveness testing equipment

cal) miles, operating a jammer against an AN/FPS-3 radar which is set to its 50 mile scale.

The essential portions of the simulation equipment are shown in the functional block diagram of Figure 14-5. Both target and jamming signals are generated at the i-f of 30 megacycles, mainly to avoid r-f simulation problems. Standard receiver and indicator units of the AN/FPS-3 radar are used; the controls are not available to the observer. The automatic programmer is controlled by a Western Electric tape transmitter and a suitably punched tape. Each time it is actuated by an answer switch, it sets the target azimuth position according to a table of random numbers. The Esterline Angus operations recorder records actual target azimuth, observer's answer as to azimuth, and experimental data. The antenna rotation simulator furnishes synchro information to the indicator, and triggers the antenna pulse generator at the proper azimuth. The antenna pulse is rectangular, and of such duration as to simulate a 3 degree beamwidth. The simulated rotation rate is 10 rpm. This is the maximum for the AN/FPS-3, and is used to minimize experiment time.

During the experiment the noise (jamming) is presented uniformly over the entire PPI. The face of the indicator is divided by lines into six pie-shaped sectors. The target is presented in the center of one of these sectors, for each of two successive antenna revolutions (scans), and the observer is required to signify his choice of the sector in which the target appeared, before the next presentation is made. The test is begun with a high S/J ratio, and proceeds by steps of one decibel to lower S/J ratios by changing noise power; the signal power is held fixed. At every step 12 presentations are made. The test continues until the observer misses more than seven targets in any step. A sequence of ten observer testings constitutes one complete jamming effectiveness experiment.

The data for each observer are corrected for the probability of correct guesses, and the result is a table of detection probability estimates for various S/J values. These are plotted on normal probability paper, as in the sample jamming curve of Figure 14-6. A straight line is fitted to the plotted data, under the assumption that the jamming curve is the normal signoid curve common in sensory preception testing. (This curve maps to a
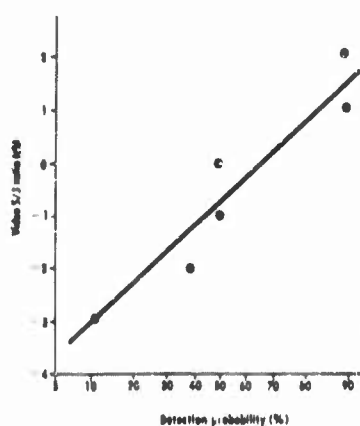
FIGURE 14-6 Sample jamming curve

straight line on probability paper.) The object is to make a maximum like-lihood estimate of the curve from which the true data arose; to do this rigorously is a complicated procedure, requiring the use of a digital com-puter. It has been found, however, that a visual estimate of a least-squares fit produces all the accuracy that is meaningful in the experiment. The 50 percent intercept is taken as the individual observer's jamming threshold. The slope of the curve, which is of the form of a standard deviation, is a measure of how slowly the observer is jammed. The jamming thresholds, in decibels, are averaged for the ten observers, to give the "mean jamming threshold", $(S/J)_{50\%}$, as the result of the experiment. The relative values of $(S/J)_{50\%}$ for different jamming sources are then the required relative effec-tiveness data. The standard deviation of the individual jamming thresholds indicates the statistical confidence level of the result. A common value for this statistic is about one decibel.

The S/J ratios have been measured at video as a matter of experimental convenience. These can be extended to i-f, where they are more meaningful, by circuit measurements. The rms error in experimental results from this program has been estimated to be less than one decibel.

Observer Selection and Training. In laboratory testing of jamming effec-tiveness it is seldom possible to use actual typical radar operators as ob-servers. Thus the observers become simulated radar operators, and the degree of realism achieved must be considered. A minimum requirement is that the results permit prediction of those which would obtain if actual radar operators were used. It would be preferable, of course, if the observers produced the same results as would the radar operators. A series of tests comparing these two groups directly (Reference 13) indicates that with reasonable care in the selection and training of observers, the desired simula-tion condition can be achieved. Following are some general recommendations on selection and training.

A set of physical requirements (dealing mainly with vision) and a fixed upper age limit are desirable, since these reduce poor performances. While consistent poor performances can perhaps be justified by the relative effective-ness argument, still they produce an undue spread in data, and thus reduce the degree of confidence in the results.

A further selection can be made on the basis of performance relative to that of the established group of observers. This selection can readily be combined with the training procedure. Training runs should be made on a jamming source which does not involve any deception; DINA is a good choice. A candidate is then accepted as an observer whenever his average performance equals or exceeds a minimum standard. This acceptance stand-

ard is set a fixed number of decibels below the average performance of the observer group on the same type of jamming. If the candidate fails to meet the acceptance standard in a given fixed number of training runs, he should be rejected for observer duty.

A final remark on observer training is that as with other skills, training must be maintained if performance is to be consistently good. If the duties are very intermittent, it may be necessary to maintain a regular training schedule for the accepted observers.

Interpretation of Data. Let us assume that the results of laboratory jamming effectiveness tests are in the form of S/J ratios at the receiver i-f output (second detector input), corresponding to a given detection probability. But the jammer designer requires a specification of the radiated pattern necessary for a certain degree of jamming, given the tactical situation and the victim's characteristics.

If we are willing to assume that the laboratory tests achieved perfect simulation, i.e., that the results are directly applicable to the field situation, then the required jamming power can be computed as follows. The antenna input ratio of signal-to-jamming per megacycle $(S_R/J_R/mc)$ is obtained from the i-f output S/J ratio by multiplying the latter by the i-f bandwidth. Then from the free-space propagation equations, the radiated jamming power required is

$$P_J = \frac{P_T \, G_T \, \sigma}{4\pi R^2 G_J \, (S_R/J_R/mc)}$$

where P_T = radar transmitted power
G_T = radar antenna gain
σ = target cross section
R = range
G_J = jamming antenna gain in the direction of the radar

If the experimental methods are similar to those described above, this value for P_J will be pessimistically (from the jammer's point of view) high. This is principally because in the laboratory, the observer's task has been made much simpler than is that of a radar operator in the field. The results of some tests performed at the University of Michigan Electronic Defense Group, give a quantitative example of this effect (Reference 14). Laboratory measurements were made, using the same jamming source, for a one-dimensional arrangement of possible target locations in the first case, and for a two-dimensional arrangement in the second case. In both cases, the targets

were presented in one quadrant of a PPI. In the first case, there were four possible locations spread uniformly in azimuth through the quadrant but all at the same range. In the second case, there were 16 possible locations, at all combinations of 4 azimuths and 4 ranges, spread uniformly through the quadrant. It was found that the threshold S/J ratio was about four decibels higher for the more difficult two-dimensional arrangement.

It is apparent, then, that accurate predictions of absolute jamming effectiveness cannot be expected from laboratory tests alone. However, a rather small amount of carefully obtained field test data can serve as an "anchoring point" for a large and detailed amount of laboratory data, thus extending the relative results into absolute results.

14.2.4 The Response of a Typical Radar Receiver to Various Types of Jamming

For purpose of analysis a search radar receiver can be assumed to consist essentially of a narrow bandpass amplifier, a detector, and a low pass amplifier. This corresponds to the i-f, detection, and video stages of a conventional receiver. Most radar receivers employ the superheterodyne principle wherein a frequency translation is performed on the incoming wave for the purpose of reducing its frequency to a more suitable value for amplification; however this process does not alter the important characteristics of the target echo and jamming ensemble. Consequently it is usually sufficient to consider the transformations occurring between the i-f amplifier input and the video amplifier output.

As a consequence of the jamming power limitation discussed in 14.2.1, the jamming waveform will not, in general, be of sufficient amplitude to cause saturation of the receiver circuits. Receiver saturation does not guarantee effective jamming; on the contrary, limiting, which is a form of saturation, is frequently utilized as an anti-jamming technique. (See 14.2.5.) For the purpose of the present analysis, however, the receiver amplifiers will be assumed to be operating in a linear region.

Calculations Using Second Order Statistics. By making use of some fundamental theorems of probability and statistics it is possible to calculate input-output relationships of various elements in the radar receiver. In general, analysis of the linear circuit elements is more straightforward; however, techniques for handling nonlinear elements, such as the detector, have been developed.

The autocorrelation function of a voltage waveform, $v(t)$, is defined as follows:

$$R(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} v(t) v(t + \tau) \, dt$$

In many problems of interest the autocorrelation function can be equivalently defined as:

$$R(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v_1 v_2 W_2(v_1, v_2, \tau) \, dv_1 \, dv_2$$

v_1 and v_2 are values of $v(t)$ at times t_1 and t_2 respectively, $\tau = t_2 - t_1$, and W_2 is the second order probability density function of $v(t)$. The autocorrelation function is closely related to the power spectrum of a waveform by the Wiener-Khinchine Theorem which states that the power spectrum is the Fourier transform of the autocorrelation function.

For linear circuits, the autocorrelation function of the output can be calculated using convolution integrals. If the circuit is described by its impulse response function, $H(t)$, then the output correlation function is:

$$R_{out}(\tau) = \int_{-\infty}^{\infty} H_c(t) R_{in}(t + \tau) \, dt$$

where $H_c(t)$ is the convolution of $H(t)$ and $H(-t)$. This leads to the input-output spectral relationship.

$$G_{out}(\omega) = |Z(\omega)|^2 G_{in}(\omega)$$

where G_{out} and G_{in} are the output and input power spectra respectively, and $Z(\omega)$ is the circuit transfer function. The reader is referred to References 15 and 16 for extensive background material and the proofs of these relationships. The latter reference also gives methods of handling nonlinear elements.

Receiver Response to DINA Jamming. For a gaussian noise input, the output of the i-f amplifier can be written as

$$v_0(t) = \alpha(t) \cos \omega_c t + \beta(t) \sin \omega_c t$$

where $\alpha(t)$ and $\beta(t)$ are independent, slowly varying time functions with gaussian properties having zero means and standard deviations σ, and $\omega_c = 2\pi f_c$ with f_c being the i-f amplifier center frequency. The voltage, $v_0(t)$, can equivalently be written as follows:

$$v_0(t) = \rho(t) \cos [\omega_o t - \phi(t)]$$

with

$$\rho(t) = \sqrt{\alpha^2(t) + \beta^2(t)} \text{ and } \phi(t) = \tan^{-1} [\beta(t)/\alpha(t)]$$

Since $\alpha(t)$ and $\beta(t)$ are independent

$$W_2(\alpha, \beta) = W_1(\alpha)W_2(\beta) = (1/2\pi\sigma^2) \exp [- (\alpha^2 + \beta^2) /2\sigma^2]$$

where W_1 and W_2 are the first and second order probability density functions.

The output of a linear (or envelope) detector will be $\rho(t)$. The first order density function of the output can be easily calculated as follows: by definition,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_2(\alpha, \beta) \, d\alpha \, d\beta = \left(\frac{1}{2\pi\sigma^2}\right) c \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left(- \frac{\alpha^2 + \beta^2}{2\sigma^2}\right) d\alpha \, d\beta = 1$$

but

$$\alpha = \rho \cos \phi \qquad \text{and} \qquad \beta = \rho \sin \phi$$

Thus,

$$\frac{1}{2\pi\sigma^2}\int_0^{2\pi} \int_0^{\infty} \exp \frac{-\rho^2(\cos^2 \phi + \sin^2 \phi)}{2\sigma^2} \, \rho \, d\rho/d\phi = 1$$

or

$$\int_0^{\infty} \frac{\rho}{\sigma^2} \exp \left(\frac{-\rho^2}{2\sigma^2}\right) d\rho = 1$$

Therefore.

$$W_1(\rho) = \begin{cases} (\rho/\sigma^2) \exp (-\rho^2/2\sigma^2), & \rho \gtrless 0 \\ 0, & \rho < 0 \end{cases}$$

This is the well-known Rayleigh distribution. The average value of the output is

$$\overline{\rho(t)} = \int_{-\infty}^{\infty} \rho W_1(\rho)\, d\rho = \int_0^{\infty} \frac{\rho^2}{\sigma^2} \exp\left(\frac{-\rho^2}{2\sigma^2}\right) d\rho = \sigma\sqrt{\frac{\pi}{2}}$$

The mean square value of the output is

$$\overline{\rho^2(t)} = \int_{-\infty}^{\infty} \rho^2 W_1(\rho)\, d\rho = \int_0^{\infty} \frac{\rho^3}{\sigma^2} \exp \frac{-\rho^2}{2\sigma^2}\, d\rho = 2\sigma^2$$

The percentage of the total video power that is essentially d-c is

$$\% \text{ d-c} = 100\, [\,\overline{\rho(t)}\,]^2/\rho^2(t) = 100\pi/4 = 78.5\%$$

Receiver Response to FM-by-Noise Jamming. In order to gain an under-standing of how a radar receiver responds to FM-by-noise jamming, it will be helpful to consider the case of a signal that is swept linearly through the receiver passband. The i-f amplifier can be approximated by a gaussian response. (A synchronously tuned amplifier approaches a gaussian response rapidly as the number of stages increases.) Thus the transfer function of the amplifier is

$$G(\omega) = A_1 \exp\,[\,-(\omega - \omega_0)^2/2b^2]$$

where A_1 is the gain at midband frequency ω_0, and b is related to the usual 3-db bandwidth β by the relation $b^2 = \beta^2/4 \ln 2$. The linearly swept signal can be written $v_i(t) = A_2 \cos(St^2/2)$ where S is the sweep rate. The ampli-fier output can be calculated by multiplying the Fourier transform of $v_i(t)$ and $G(\omega)$, then taking the inverse transform of the result. The envelope of the output $v_o(t)$ is found to be

$$v_o(t) = \frac{A_1 A_2}{\sqrt[4]{1 - (S/b)^2}} \exp\left\{\frac{-S^2}{2b^2[1 + (S/b^2)^2]}\left(t - \frac{\omega_0}{S}\right)^2\right\}$$

or

$$v_o(t) = \frac{A_1 A_2}{\sqrt[4]{1 + a^2}} \exp \frac{-(t - t_0)^2}{2(1 + a^2)/aS}$$

where $a = S/b^2$ and $t_0 = \omega_0/S$.

Consider first the case of slow sweep. If $a \ll 1$, then

$$v_c(t) \sim A_1 A_2 \exp \frac{-(t - t_0)^2}{2/aS} = A \exp \frac{-(t - t_0)^2}{2(b/S)^2}$$

with $A = A_1 A_2$. This case for a given bandwidth and several sweep speeds is illustrated in Figure 14-7. The receiver output is a pulse of constant
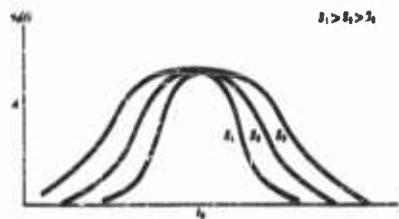
FIGURE 14-7 Response of a receiver to a linearly swept signal (slow sweep case)
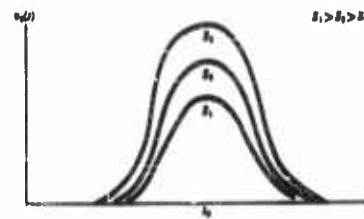
FIGURE 14-8 Response of a receiver to a linearly swept signal (fast sweep case)

amplitude whose width varies directly with receiver bandwidth and inversely with sweep rate. In the case of fast sweep rates, i.e., $a \gg 1$,

$$v_o(t) \sim \frac{A_1 A_2}{\sqrt{a}} \exp \frac{-(t - t_0)^2}{2/aS} = Ab/\sqrt{S} \exp \frac{-(t - t_0)^2}{2(1/b)^2}$$

Here the situation is quite different. As shown in Figure 14-8 the output pulse amplitude is directly proportional to the receiver bandwidth and decreases with increasing sweep speed. The pulsewidth, however, remains essentially unchanged as the sweep speed changes. It depends only on the reciprocal of the receiver bandwidth.

By utilizing the foregoing considerations the response of a receiver to FM-by-noise can be explained. Assuming a barrage width that is large compared to the receiver bandwidth, the sweep speed during any particular transit of the receiver passband by the randomly sweeping jamming will depend, essentially, on jamming barrage width and the spectral composition of the modulating noise. If the noise spectrum has a sufficiently low upper cutoff frequency (FM-by-LF noise), the receiver output will consist of a series of impulses of random amplitude and random structure. As the modulating noise frequency is increased (FM-by-WB noise), the random impulses will begin to overlap since the pulsewidth depends only on the receiver bandwidth. As a direct consequence of the Central Limit Theorem of probability theory the statistics of the noise so produced at the i-f amplifier will be essentially gaussian.

Because of the gaussian properties of the jamming waveform that is produced in the receiver by FM-by-WB noise, calculations of the a-c and d-c

power output are identical to those given for DINA. In order to meet the requirements for producing jamming that is equivalent to DINA, the following inequality must be satisfied:

$$f_R < f_N < f_J$$

where f_R is the receiver bandwidth, f_N is the average noise bandwidth, and f_J is the jamming barrage width. Typical numerical values for these qualities are $f_R = 1$ mc, $f_N = 5$ mc, $f_J = 200$ mc.

The receiver output power for FM-by-LF noise jamming can be calculated to good approximation using the quasi steady state method of analysis. The receiver is assumed to have a gaussian frequency response, thus

$$v_o(\omega) = A_1 \exp \frac{-2 \ln 2(\omega - \omega_0)^2}{\beta}$$

where β is the usual 3-db bandwidth and $\omega_0 = 2\pi f_0$, f_0 being the receiver center frequency. The input voltage is

$$v_i(t) = A_0 \cos \left[\omega_0 t + \int' x(t') \, dt' \right]$$

where ω_0 is the carrier frequency and $x(t)$ is a function of the modulating noise which has gaussian statistics. For slow sweeps, we assume that $x(t)$ is a constant function of time. For the case where ω_0 and ω_0 coincide, i.e. the jammer is tuned to the receiver center frequency, the detector output is approximately

$$v_o(t) = A_0 A_1 \exp \frac{-2 \ln 2 x^2}{\beta^2}$$

Using the gaussian properties of $x(t)$, we can calculate the average and mean square value of the output as follows:

$$v_o(t) = \frac{A_0 A_1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp \frac{-2 \ln 2 x^2}{\beta^2} \exp \frac{-x^2}{2\sigma^2} \, dx = \frac{A_0 A_1}{4\sigma^2 \ln 2 + \beta^2}$$

$$\overline{v_o(t)} = \frac{A_0^2 A_1^2}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp \frac{-4 \ln 2 x^2}{\beta^2} \exp \frac{-x^2}{2\sigma^2} \, dx = \frac{A_0^2 A_1^0 \beta}{\sqrt{8\sigma^2 \ln 2 + \beta^2}}$$

Note: σ is the rms frequency of the jamming waveform.

The percentage of total power contained in the d-c component of the output is

$$\%\text{d-c} = \frac{[\overline{v_a(t)}]^2}{\overline{v_a{}^2(t)}} = \frac{\beta\sqrt{8\sigma^2 \ln 2 + \beta^2}}{4\sigma^2 \ln 2 + \beta^2} \sim 85\,\frac{\beta}{\sigma} \text{ for } \sigma >> \beta$$

For typical values of bandwidth β and deviation σ this indicates that the majority of the video jamming power is contained in the a-c component. This superiority of FM-by-LF noise over FM-by-WB noise or DINA in producing effective jamming power at video has been verified experimentally; however, as was suggested in 14.2.2 and as will be discussed further in 14.2.5, there exist counter-countermeasures that alter this observed superiority considerably.

In the previous analysis of FM-by-noise jamming the jammer center frequency and the receiver band center were assumed to coincide. For successful barrage jamming it is necessary that the jamming be uniformly effective throughout the barrage width.

The first consideration is the production of uniform jamming power across the barrage. Since the voltage amplitude distribution of the modulating noise determines the characteristics of the frequency excursions of the jamming signal, this parameter also determines the actual power-frequency spectrum of the jamming barrage. If the modulating noise has gaussian statistics the resulting power spectrum will have the characteristic bell-like shape of the gaussian function. In order to produce a uniform power spectrum it is necessary to alter the voltage amplitude distribution of the modulating noise before it is used to modulate the carrier. Middleton has shown that the necessary transformation is the error function (Reference 5).

Starting with the modulating voltage $x(t)$ it is desired to find a transformation g such that $y(t) = g\,[x(t)]$ has a first order probability density $W_1(y)$ that is constant over the interval from $y = -a$ to $y = +a$. In order to satisfy the requirement that

$$\int_{-\infty}^{\infty} W_1(y)\, dy = 1$$

$$W_1(y) = \begin{cases} \dfrac{1}{2a} & \text{for } -a < y < a \\[2mm] 0 & \text{for } |y| > a \end{cases}$$

Now

$$\frac{1}{2a}\int_{-\infty}^{y} dy' = \frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{x} \exp\frac{-x'^2}{2\sigma^2}\, dx$$

$$\frac{1}{2} + \frac{1}{2a}\int_0^y dy' = \frac{1}{2} + \frac{1}{\sqrt{2\pi}\sigma}\int_0^x \exp\frac{-x'^2}{2\sigma^2} dx'$$

$$y = \sqrt{\frac{2}{\pi}} \frac{a}{\sigma}\int_0^x \exp\frac{-x'^2}{2\sigma^2} dx'$$

Let

$$\frac{x'^2}{2\sigma^2} = Z^2$$

then

$$y = \frac{2a}{\sqrt{\pi}}\int_0^{x/\sqrt{2}\sigma} e^{-z^2} dz = a \, \text{erf}\left(\frac{x}{\sqrt{2}\sigma}\right)$$

Thus the desired transformation g, is the error function. By symmetry it is seen to hold for the region $y < 0$ and $x < 0$. A device that approximates the error function (called an "erfer") has been constructed and successfully used to produce a uniform jamming barrage (Reference 17).

Another consideration affecting the uniformity of jamming effectiveness in the barrage is the time structure of the jamming waveform. This will depend on the distribution of traversals of the receiver passband, or equivalently the distribution of level crossings at various amplitudes of the modulating noise. "Erfed" noise does not produce uniform level crossings throughout the barrage, but tends to fall off near the ends. If the modulating noise bandwidth sufficiently exceeds the receiver bandwidth there will still be sufficient crossings to produce gaussian noise at the receiver i-f output. Modulator design limitations may preclude the use of an "erfer" noise modulation having sufficient bandwidth to produce an adequate radar receiver traversal rate near the ends of the barrage, particularly when the victim receiver employs a very wideband Dicke Fix. An alternative modulation consisting of the sum of a high-frequency periodic wave and noise may be used to produce a jamming barrage having a relatively uniform power spectrum and receiver traversal rate throughout the entire barrage width (Reference 5).

14.2.5 The Influence of Anti-Jamming Techniques on Jamming Effectiveness

In the final assessment of the effectiveness of a particular jamming modulation against a search radar system, it is necessary to concede that the victim radar possesses and makes efficient use of any appropriate anti-jamming (A-J) techniques. It is conceivable that a jamming technique that appears to be quite effective against a conventional radar may be rendered useless

by the utilization of a simple A-J device. Consequently, it is of interest to review some of the existing and proposed A-J techniques. For a more complete listing, the reader is referred to Reference 18.

The ultimate goal in radar design is to produce equipment that will continue to perform its specified function regardless of the prevailing jamming environment. Because of physical considerations, absolute invulnerability can not, in general, be achieved. However, insofar as is practicable, such A-J devices as are required to optimize radar performance in the expected jamming environment should be provided. These A-J features may take the form of (1) devices that give the radar maximum probability of detecting a jammer carrying target, or (2) devices that prevent the radar being saturated or overloaded with false alarms so that, except for detecting the jammer carrying target, the radar can continue to function effectively. The latter class of devices are generally referred to as Constant False Alarm Rate (CFAR) techniques.

An A-J technique may consist of a relatively simple alteration or addition to an existing radar or it may result in an entirely new radar system design. Many of the useful A-J devices are installed in a radar system in a manner such that they may be selected for use at the discretion of the operator. This gives the radar system the flexibility to allow adjustment for optimum performance according to the existing jamming situation.

A-J Techniques Related to Frequency Characteristics. Search radar frequency agility, the importance of which was emphasized in 14.2.1, may be considered as an A-J technique. Frequency agility may consist simply of the capability of fast mechanical tuning or, in the ultimate form, a capability of changing frequency on a pulse-to-pulse basis. A related technique, called frequency diversity, consists of choosing the frequencies of the individual radars in a radar network so that they are distributed throughout the entire spectrum of frequencies that are suitable for search radars. The combination of frequency diversity with frequency agility causes maximum spreading of the available jamming energy necessitating the dilution of a jamming effort in order to cover the search radar frequency band.

Another counter-countermeasure (CCM) related to the radar's frequency characteristics is called diplexing. It consists of using two transmitters which transmit pulses that are separated in frequency and time using the same antenna. The returns are separated using two frequency selective channels and the signal in the channel which contains the leading pulse is delayed such that the echoes now coincide in time. (Note: In some installations, time separation and subsequent delay are not utilized.) Video data from the two channels are fed into a two input coincidence gate, whose output is always the

minimum of its inputs. Thus the radar is impervious to jamming that covers only one of the two transmitter frequencies. The obvious counter-counter-countermeasure (CCCM) against diplexing, as well as against frequency diversity and frequency agility, is the simultaneous jamming of all radar frequency bands in which activity is detected.

A-J Techniques and FM-by-Noise Jamming. The desirability of using frequency modulation jamming is a direct result of the fact that electronically tunable microwave power generators readily provide a broadband jamming source that can be rapidly tuned to the frequency range of interest. The mechanism by which an FM source produces jamming power in what is essentially an AM receiver was discussed in 14.2.4. One A-J technique attempts to reduce the effectiveness of FM jamming by effectively reducing the response of the radar receiver when the jammer sweeps through the receiver passband. Other A-J schemes operate on the time structure of the jamming waveform and seek to discriminate in favor of the target returns.

Devices that make use of the former technique are variously referred to as FM jamming blankers, guard band receivers, or "rabbit traps". One realization of such a device is shown in a simplified block diagram in Figure 14-9. The response of the wideband amplifier is fed to two paths, (1)
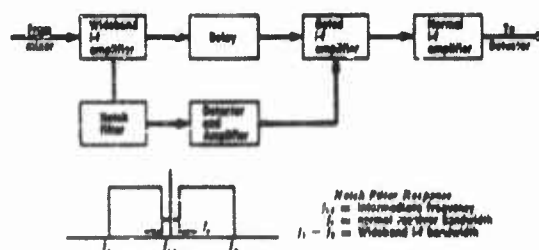
Figure 14-9　Block diagram of an FM jamming blanker

through a narrow-band notch filter centered at the i-f frequency and then through a detector and amplifier, (ultimately the video pulse is applied to the gated i-f amplifier); and (2) through a delay line and thence to the input of the gated i-f amplifier. The delay is adjusted so that the characteristic delay of the circuitry in the other path is balanced, and the i-f signal and the video gating pulse arrive at the gated amplifier simultaneously. Thus the gated i-f amplifier cuts off the receiver when a jamming signal sweeps through the receiver passband. The notched filter rejects the narrow-band target return thus preventing operation of the gating circuit by a desired signal; therefore the target echo will negotiate the receiver and be displayed

except when it occurs simultaneously with an impulse produced by the jammer. This technique is quite effective against FM-by-LF noise. Its effectiveness decreases as the modulating noise upper cutoff frequency is increased. When the condition for overlapping impulses at the wideband amplifier output is reached the device is gated off almost constantly, and consequently is no longer effective.

Another A-J device, which has found wide application as a search radar CCM technique, is known as the Dicke Fix. This device consists essentially of a wideband i-f amplifier that is inserted in a conventional radar receiver between the mixer and normal i-f amplifier as shown in Figure 14-10. The
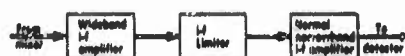
FIGURE 14-10   The Dicke Fix

action of the Dicke Fix against FM-by-noise jamming can be qualitatively described as follows: The impulsive responses of the wideband i-f amplifier to the FM jamming have a large amplitude with respect to the target echo. The i-f limiter limits the jamming pulses to a level commensurate with the echo. These clipped jamming pulses are spread over a broad spectrum, so that after the narrow i-f filtering they are small compared to the echo. Also, the target returns are coherent on successive radar sweeps whereas the random pulses are not; therefore the target pulses add constructively during the integration produced by the display and thereby can be seen above the noncoherent jamming pulses. The Dicke Fix is quite effective against FM-by-LF noise, but like the FM blanker decreases in effectiveness as the modulating noise upper cutoff frequency is increased. Higher modulating noise frequencies are required to produce overlapping impulses (and consequently gaussian noise) in the case of the Dicke Fix as compared to normal receiver because of the narrower impulse response with the wideband i-f amplifier. The use of multiple independent FM-by-noise jammers produces an effect similar to that of increasing the modulating noise bandwidth of a single jammer, in that the average number of traversals of the victim receiver is increased in both cases. In addition to its usefulness against FM-by-LF noise, the Dicke Fix is also quite effective against other types of interference which have a characteristic pulse-like time structure. An extensive account of both the theoretical and experimental studies of the Dicke Fix and other important A-J techniques is contained in Reference 19.

Constant False Alarm Rate (CFAR) Techniques. The CFAR devices have been developed to prevent overloading or saturation by false alarms caused by a strong jammer. This is especially necessary in the case of an automatic detector where a target is reported whenever a preselected threshold level is exceeded. The CFAR device causes the radar to respond to a

signal to noise ratio instead of an absolute signal level above a fixed threshold.

One group of CFAR devices operates on the principle of automatically adjusting the receiver gain to maintain a constant output. The gain control time constant must be long compared to the target pulsewidth; consequently this technique is most effective against jamming that produces long pulses in the radar receiver.

A more effective method of achieving CFAR performance results from the use of the zero-crossing statistics of the signal and noise. The Dicke Fix circuitry can be used along these lines and as such is called Dicke Fix CFAR. For CFAR operation the limit level of the i-f limiter is set at a level that is well down into the receiver noise. Thus any increase in the receiver input does not effect the output. The limiting does not destroy the zero-crossing information; consequently a large signal to noise ratio will result in a detectable target. As the signal to noise ratio decreases, the target zero-crossing information becomes submerged in the noise, resulting in the disappearance of the target indication. By a fortuitous combination of circumstances, noise alone can produce a false target indication; however, by the proper adjustment of receiver parameters, the false alarm rate can be maintained at a constant and predictably low level.

Combined Passive-Active Detection Systems. In the event that a radar is successfully jammed in range but still retains the capability of passively determining target azimuth, there exists the possibility of utilizing similar information from radars at other geographical locations to fix the target location by triangulation. The principal difficulty encountered with such a system occurs in a multiple jammer environment. In this case an intersection of azimuth strobes may or may not correspond to an actual target depending on whether or not the strobes in question were generated by the same jamming source. Several systems, which utilize correlation techniques to eliminate ghost intersections, have been proposed and are under development (Reference 20).

Successful jamming of passive-active systems that utilize correlation techniques presents a difficult problem, principally because a correlation device performs best against a random signal. Several methods of jamming (not without considerable practical difficulties) have been suggested (Reference 21). For example, transmitting uncorrelated jamming signals on different azimuth from the jamming vehicle would appear to be effective from theoretical considerations; however it is difficult to avoid having sufficient power radiated from the back of the jammer antennas to allow successful operation of the correlators. Another CCCM, which consists of transmitting correlated

noise jamming from widely separated sources, is complicated by the difficulty of maintainin correlation between two high frequency random noise sources.

New Concepts in Search Radar Design.   During recent years search radar systems have been proposed (several are in varying stages of development) that differ somewhat radically from the conventional rectangular pulse radars. The principal difference exists in the characteristics of the transmitted signal. Rectangular pulses have found wide usage in radar because of the obvious ease with which range resolution can be obtained and also because of the predominance of microwave power sources with high peak power but low average power capability. The advent of microwave sources, traveling wave tubes, that operate most efficiently when the difference between the peak and average power is minimized, has caused attention to be focused on the problem of achieving the benefits of more or less continuous wave transmission without loss of range resolution.

Essentially this problem is solved by transmitting a relatively complicated waveform, then comparing the echo returns to a stored replica of the transmitted wave to establish the epochs of the various echo producing targets. The details of several systems are contained in References 22, 23, and 24. While it is not claimed that these systems will be jam-proof, it is true that an entirely different jamming philosophy (as compared to conventional search radar jamming) may be required in order to achieve optimum jamming effectiveness.

14.2.6 Some Typical Results

The following jamming effectiveness results of the Johns Hopkins program are excerpts from Reference 12. They describe the performance of important barrage jamming sources against a typical search radar which is unprotected by ECCM. Detailed coverage of the test conditions will not be attempted here.

The victim radar is the AN/FPS-3. The experimental method generally follows the recommendations of Section 14.2.3. The units of the FPS-3 which are either used or simulated in the program are operated with normal parameter values for that radar.

The curves of Figure 14-11 result from circuit measurements on the FPS-3 normal receiver, and relate S/J ratios at video output (CRT grid) to those at i-f output (second detector input) for several important barrage jamming sources. The dots on the curves are typical values for the mean jamming threshold, $(S/J)_{50\%}$. The S/J ratios corresponding to detection probabilities other than 50% may be located on these curves by use of Figure 14-12, which is the approximate jamming curve for the "average"
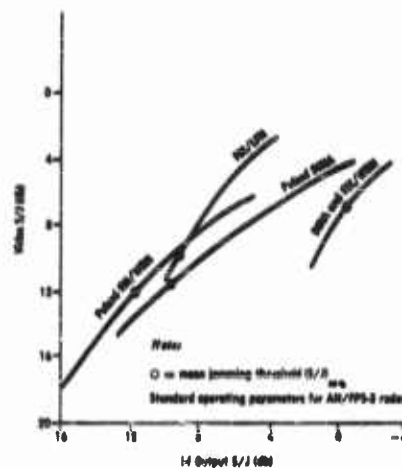
FIGURE 14-11   Video S/J versus i-f output   FIGURE 14-12   An S/J extrapolation curve
S/J for various jamming sources   for detection probabilities other than 50%

observer. The curves and dots of Figure 14-11 are representative of the jam-
ming sources listed, with the following restrictions on the jamming para-
meters:

DINA   . . . . barrage width must be several times greater than victim
                    receiver bandwidth. Clipping of the DINA does not change
                    its effectiveness.

FM/LFN . . . . Upper cutoff frequency of modulating noise must be below
                    50 Kc. Deviation must be several times greater than victim
                    receiver bandwidth.

FM/WBN . . . Upper cutoff frequency of modulating noise must be above
                    500 Kc. Deviation must be several times greater than vic-
                    tim receiver bandwidth.

RPB Sources   Pulse recurrence period $= 80 \mu s \pm 30\%$
              Pulse duration $= 4 \mu s \pm 30\%$
              Ratio of pulse peak to background (voltage) $= 4 \pm 33\%$
              ($\pm$ term is the peak random variation from the nominal
              value.)

For the FM sources, the jamming effectiveness is not changed by using
"erfed modulating noise to "whiten" the jamming spectrum, as suggested
in 14.2.4.

Within the above restrictions, the jamming effectiveness of these sources

is constant over wide ranges of such parameters as deviation, upper and lower cutoff frequencies of modulating noise, location of the victim in the barrage, etc. The pulse parameters given are those used in limited testing of RPB jamming; limiting values are not known.

From Figure 14-11 it is seen that the pulse-like sources are 10 to 12 decibels more effective than the continuous sources. Nevertheless, FM-by-LF noise is of questionable value, since its intermittent time structure makes ECCM quite easy. The pulsed sources offer high jamming effectiveness without being appreciably easier to counter than are their unpulsed counterparts. For example, with the pulse parameters listed, about half of the average jamming power is in the pulses, and the other half is in the background jamming between pulses. So even if the effect of the pulses were completely nullified by ECCM, the loss due to investing jamming power in these pulses would be only 3 db. Against this we have a 10 to 12 db gain in jamming effectiveness if no ECCM is used.

It is noteworthy in Figure 14-11 that the variation in $(S/J)_{50\%}$ is much greater at i-f than at video. This indicates that there are greater differences in circuit effects (Section 14.2.4) than in psychophysical effects.

14.3 The Tracking Radar Jamming Problem

14.3.1 Statement of the Problem

The term "Tracking Radar" as used here includes several different types of radar, all of which have as their primary function to automatically supply position data on one or more selected targets for weapon control purposes. In most applications the coordinates supplied by the radar are azimuth, elevation and slant range, although in some cases radial velocity may be measured instead of, or in addition to, range.

Angle data is usually determined by using servo loops to keep the radar antenna pattern centered on a selected target in which case only a single target can be tracked by each radar. In some more recent types multiple target capability and/or ability to continue searching while tracking are achieved by using a multiple beam antenna pattern or by time sharing a single beam. While this discussion will be concerned primarily with the single target class, many of the comments are applicable also to the multiple target class.

Range information, when desired, is obtained as in search radars by measuring the time delay of the target echo, so pulsed radar is used when precise target range is required.

Target velocity information can be obtained from the time derivatives of the above coordinates, but if only the radial component of velocity is re-

quired it is often obtained directly from the doppler shift of the echo signal relative to the transmitted signal.

An important part of the tracking process is the necessity of discriminating against radar returns from targets other than the selected ones. Discrimination in angle is accomplished by the antenna pattern and in range by time selection or gating of the received signals. In the case of doppler radars target signal selection is achieved on the basis of radial velocity. This interdependence between tracking channels may, under certain circumstances, become an important factor in ECM considerations.

The object of ECM against radars is of course to degrade the target position data supplied to a weapon by an enemy radar to the extent that the probability of success of the weapon is at least decreased (Reference 25). The application of a tracking radar, particularly the type of weapon it is controlling, will strongly influence the relative importance of the various tracking channels (i.e. angle, range and/or radial velocity) and will also be the most important factor in determining the extent to which the tracking data must be degraded for the ECM to be considered successful. A thorough treatment of this aspect of the problem will not be attempted here but some of the more obvious points will be mentioned briefly.

If a radar is to supply target position data for aiming of guns or rocket launchers or for command guidance of a missile, a higher degree of accuracy is required than if the data is used only for steering a missile to the vicinity of a selected target to allow a homing radar to lock onto the target. Similarly, a given angular error would have a much more serious effect on the radar supplying command guidance information than it would on the homing radar, since the latter is much closer to the target and the range is constantly decreasing. It is also important to note that in some cases, for instance in a homing device using proportional navigation, angle rate information determines the steering commands rather than angle information itself.

Another important factor which depends upon the type and application of the victim radar is the extent to which an operator can assist or take control of the target tracking and the time required for reacquisition of a target if lock-on is broken. This will be an important consideration in determining the effectiveness of deception jamming and of intermittent noise jamming.

For example in the case of a ground-based tracking radar an operator can give his full attention to monitoring the range tracking function and, if necessary to reject intermittent or deceptive jamming, can assume full control of the range gate. At the other extreme in this respect is the airborne interceptor radar which is operated by the pilot who must monitor the other tracking channels as well as operate any A-J devices and control the aircraft.

The tactical situation in which the jammer is used will also influence the

effectiveness of certain types of jamming signals insofar as it will determine the characteristics of the radars to be encountered. If violent maneuvers are to be expected, the tracking circuits of the radar must have much faster response capabilities and hence the time required to introduce an appreciable error will be greatly reduced.

As in the case of search radars, jamming of tracking radars can be roughly divided into two groups, obscuration and deception, depending on the intended effect on the victim radar, although in the case of tracking radars there is perhaps more overlapping of the two groups.

The desired effect of obscuration is to completely destroy, or at least increase the uncertainty of, target position data while deception jamming causes the radar to supply possibly smooth but erroneous target range, angle and/or velocity data. Specific examples of these two types of jamming depend upon the tracking channel to be jammed and the type of radar.

In general deception jamming requires much less average power than does obscuration, but necessitates a more detailed knowledge of the parameters of the victim radar and usually requires more nearly continuous reception of the victim radar's transmitted signal. The last requirement gives rise to fairly severe look-through and antenna isolation problems. The difficulty of achieving multiple radar jamming capability is perhaps the principal weakness of this type of jamming.

14.3.2 Influence of Tracking Type on Jamming Signals

14.3.2.1 Range Trackers

The effectiveness of jamming signals against range trackers is quite important since the results are applicable to nearly all pulse tracking systems, including those that use monopulse techniques for angle tracking as well as some track-while-scan systems. It is interesting to note that when monopulse or certain types of conical-scan angle tracking are used, the range tracking loop is probably the most vulnerable part of the system.

The operation of a split-gate range tracker can be briefly described as follows: As shown in Figure 14-13, two adjoining gate pulses, each having a time duration approximately equal to that of the radar pulse, are generated within the radar. The time delay of the pair of gate pulses, relative to the transmitted pulse, is made to correspond roughly with that of the desired target pulse, either manually or by an automatic search circuit.

The gating pulses are used separately to gate the receiver video signal, and the integrated signal contents of the two gates are then compared. An error voltage is developed which is proportional to the difference in signal content. This error voltage is used to adjust the position of the pair of gates until their signal contents are equal. The error voltage is integrated once or twice
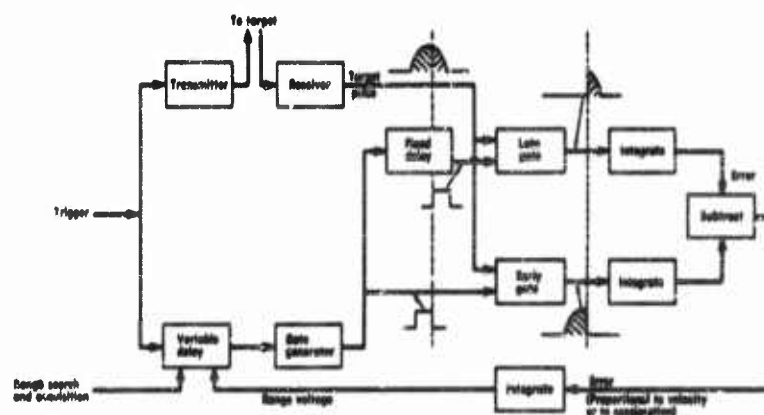
FIGURE 14-13 Typical automatic range tracker

before it is used to adjust the delay of the gate pulses. The characteristics of this integrating circuit are one of the primary factors in determining the speed of response of the tracking circuits to target motion and to a jamming signal.

Tracking loops which have a single integration in the feedback path are referred to as velocity coupled since the error, being integrated once to give a position correction, corresponds to the velocity of the gates. If no error is detected the velocity goes to zero and the position of the gates remains constant. If two integrations are performed on the error to obtain the correction voltage, the loop is said to be acceleration coupled. If the error voltage is zero, the acceleration of the gate position is zero and the velocity remains constant.

In a completely automatic tracking system the integrating time constants are limited by the required dynamic capabilities of the system, so that in effect the maximum "position memory" or "velocity memory" of the system is limited. It is obvious though that both of these features, particularly the latter, are of considerable value to the tracking system when subjected to intermittent jamming which will obscure the target for short lengths of time.

If, for any reason, the target pulse and range gates move far enough apart so that the gates no longer coincide with any part of the target pulse, then the system is no longer locked on the target and reacquisition is necessary.

Deception. If average power requirements alone are considered, by far the most efficient jamming signal against the split gate range tracker is the gate-sealer or gate-grabber.

Since the tracker centers the range gate on the "center of gravity" of the time-selected echo pulse, it is obvious that shifting the apparent center of gravity of the signal away from the true target position will cause an error in the range output data. In the gate-grabber jammer a series of pulses, similar to the target pulses and somewhat larger, is generated and synchronized with the radar pulses so that each jammer pulse coincides in time with the true signal return within the radar. The position of these pulses in time is slowly changed so that they move away from the true target and cause the apparent center of gravity to shift, so that the range gates move away from the true target. This process is continued until the range gates no longer contain the target pulse and the false pulse is then turned off, causing break-lock. The jammers which accomplish this are described in Chapter 15.

If the apparent acceleration of the false target is within the response capability of the radar, the average jamming power merely has to exceed the average skin return power in order to be successful.

It is clear that a high degree of synchronization between the jammer and the radar is required for this type of jamming. The repeater jammers described in Chapter 15 provide this synchronization very effectively, but since transmission and reception occur simultaneously and since a target with a large effective cross section requires a high gain repeater, isolation between the receiving and transmitting antennas can be a serious problem.

To overcome the isolation problem a time sharing technique is sometimes used which involves alternate gating off of the repeater receiver and transmitter so that regeneration is prevented. The retransmitted signal from this type of repeater consists of a burst of pulses, each much shorter than the radar pulse. The radar receiver bandwidth being designed for the wider pulse will not be able to resolve the fine structure of the repeated signal and the burst will appear as one pulse. Of course a radar receiver can be modified (References 26 and 27) so as to be able to discriminate against this type of signal, but this involves increasing the bandwidth and possibly increasing the susceptibility to noise jamming.

Several other anti-jam schemes have been proposed to counter repeater jamming. Since apparent motion of the signal center of gravity is caused by increasing the delay of the retransmitted pulse, the pull-off can only be accomplished toward increasing ranges. If the range tracker is made to track only the leading edge (or the leading portion) of the signal pulse it will ignore the false signal, unless it is much larger than the true signal return. Moreover, if the portion of the signal pulse sampled by the range tracker is narrower than the minimum inherent delay of the repeater, the false signal pulse will be completely ignored by the tracker. Pull-off can be accomplished in the direction of decreasing range by delaying the jammer pulse by a little

less than a repetition period but this move can be countered by varying the repetition rate of the radar. If pull-off is achieved a rapid search toward decreasing range as soon as the false target disappears will minimize the off target time.

Limiting the acceleration capability of the tracker to values which can reasonably be expected from a target also decreases the vulnerability of a radar to this countermeasure.

A combination of the above counter-countermeasures would at least place very severe requirements on a gate-grabber type jammer but would result in considerable added complexity of the radar.

A similar jamming technique has been proposed which, while requiring somewhat higher average power than does a repeater, should be considerably less vulnerable to countercountermeasures (References 28, 29, and 30).

This technique consists of generating a train of pulses, at the appropriate r-f frequency and pulsewidth, having a prf very near but not equal to an integral multiple of that of the victim radar. Such a jamming signal will cause at the range tracking circuits a series of pulses which appear to be moving slowly past the actual signal pulse, and the tracking gates will be pulled off the true target pulse in much the same manner as in the case of a repeater. However, since the jamming pulses pass through the leading edge of the true target, edge tracking should afford much less protection for the radar. Even though limited acceleration capability of the tracker may prevent the range gates from actually locking on and tracking a jamming pulse, the cumulative effect of a series of false pulses can move the gates completely off the true target. Reacquisition of the target pulse, particularly in an unmanned radar, should be very difficult with this type of jamming.

If the victim radar is using manually-aided range tracking it is quite possible that the judgement of a human operator and his knowledge of the tactical situation would enable him to completely reject the false targets generated by either of the above jamming techniques.

Obscuration—Noise Jamming. In this section are included those types of jamming signals which are random in character and are essentially unsynchronized with the victim radar, the object of the jamming being to overwhelm the target signal with another signal containing no information regarding the range of the target (References 31, 32, 33, and 34).

The effect of noise jamming on a range tracking radar can be considered in two parts—the effect of the radar receiver itself on the received jamming signal and the effect of the resulting video in producing erroneous tracking data.

For the present let it be assumed that, regardless of the characteristics of the jamming signal at the input of the receiver, the local time structure of

the resulting video will be quite similar to that of the true pulse because of the narrow filter through which they both have passed. That is, the video will roughly resemble a random sequence of target pulses, the amplitude and spacing of which will of course depend on the original jamming signal at the receiver input, but will be random.

As the split-gate tracking circuits sample this jamming video along wtih the signal pulse, there will be, on some samples at least, a jamming pulse appearing within the range gates but not coinciding with the center of the gate pair. This will cause a false error signal resulting in a change in position of the tracking gates, which will not be centered on the true target. A sequence of such samplings results in the range gate being jittered about the true target position, the magnitude of the jitter depending on the amplitude of the jamming pulses, their frequency of occurrence, and the response of the tracking circuits to a given apparent error. There is of course a limit to the amplitude of jamming pulses which the receiver will pass and the response of the tracking circuits is such that a single jamming pulse will not cause sufficient error to move the tracking gates completely off the target pulse, but as the jamming pulses become large and/or more frequent the probability becomes appreciable that a large enough succession of errors will occur in the same direction to move the gates off the target pulse, and break-lock occurs.

From the above considerations it appears that the rms error caused by a given jamming signal would vary inversely as the tracking time constant and this is indeed the case. However, break-lock is a peak error phenomenon and it can be shown (Reference 31) that the J/S ratio required to break-lock is essentially independent of the response time of the tracking circuits. The time required to break-lock however, once the required J/S ratio is achieved, is proportional to the tracker response time so that long tracker integrating times (i.e. a slowly acting servo) affords a great deal of protection against intermittent jamming. Of course, as implied above, a long integrating time also makes successful gate grabbing more difficult.

It was stated in the discussion of the effect of noise on the tracking circuits that both the amplitude and frequency of occurrence of the jamming pulses influence the resulting error. It follows then that large, scarce pulses can be as effective as smaller frequent pulses. However, it is a simple matter for the radar to be made to discriminate against these large, scarce pulses merely by limiting the video at the peak value of the signal.

If the jamming signal is such that the jamming pulses at the radar video are overlapping to a large extent, limiting alone no longer affords protection to the tracking circuits since the target pulse will now be riding on top of the jamming signal and will be destroyed by the clipping.

In considering jamming of a range tracking radar it should be kept in

mind that large, ground based tracking systems can afford to and often do have a human operator either monitoring the range channel or manually tracking the target using an A-scope presentation.

Considerable work has been done on the effects of noise and jamming on an A-scope presentation (References 1, 35, 36, and 37). Lawson and Uhlenbeck particularly have considered the role of various radar parameters on these effects. Many of the factors considered in these studies are of interest only in the jamming of an A-scope in a search radar.

In the case of tracking radar, where the target is being searchlighted and the target pulse is constantly displayed on the scope, successful noise jamming requires that there be essentially no blank spaces in the jamming video. If even a fairly small percentage of the target pulses are unobscured, the correlating effect of the A-scope allows them to be seen through the noise. If frequent but short blank spaces occur in the jamming, the baseline of the presentation becomes pronounced and the target pulse can be detected by the "baseline notch" effect.

In view of the above discussion, it appears that a criterion for effective, not easily countered, obscuration jamming of either an automatic range tracker or a manually aided tracker is that the video noise produced by the jamming be dense in the vicinity of the target pulse for essentially every repetition of the pulse (Reference 38). Assuming no synchronization between the jamming and the radar, this means that the jamming video be dense all of the time (i.e. no holes in a time plot of the video).

It should be kept in mind that, unless the spectrum of the jamming video covers the same band as does that of the signal, the jamming can be at least partly discriminated against with simple filters.

With the above considerations in mind, the effectiveness of various jamming signals against range trackers can be considered. The techniques for generating these signals have been discussed in other chapters. The choice between DINA, AM-by-noise or FM-by-noise for a particular jamming application would probably depend more on such considerations as ease and efficiency of generation and the width of the r-f spectrum to be covered than on the effectiveness of the resulting signal.

The effect of DINA on an automatic or manually aided range tracker is for most purposes the same as receiver noise if the width of the transmitted spectrum is greater than that of the victim radar. The latter condition should be satisfied to insure the generation of sufficiently wideband video noise.

Amplitude-modulation by noise is easily generated at relatively high power levels and for this reason is often used in applications where a fixed or slowly tunable spot jammer can be used, that is against fixed frequency radars.

The development of voltage tunable power tubes has also made this type

of jamming practical against even rapidly tuned radars with the help of some sort of automatic frequency lock-on technique. Here again though, the jammer output spectrum must be at least as wide as the passband of the victim radar in order to insure sufficiently high frequencies in the resulting video. This means of course that the jammer modulating noise would have a spectral distribution roughly equivalent to the passband of the victim radar's video amplifiers. This may be a rather severe requirement against radars using very short pulses on the order of 0.1 μsec.

The availability of high power BWO's such as the Carcinotron has made the generation of FM-by-noise practical at power levels required for jamming. This type of jamming is particularly suited to barrage jamming over a wide band of frequencies, since the width of the output spectrum is determined by the amplitude of the modulation signal rather than by its bandwidth.

The barrage width required from a jammer will probably be dictated by such considerations as the number of jammers available to cover a given barrage. Since the effectiveness of FM jamming against an AM receiver depends on slope detection on the skirts of the radar receiver response curve, the jammer frequency deviation should be at least several times as wide as the victim radar bandwidth, but since a need for wide frequency coverage is assumed, this requirement will automatically be satisfied.

The response of a radar receiver to FM-by-noise has been discussed earlier in this chapter and it will suffice to mention here that the need for producing wideband, everywhere dense video noise in the radar receiver requires that the band of noise used for modulating the jammer be at least as wide as the victim receiver and preferably twice as wide. In effect what is required is that the average time interval between passes of the jammer carrier through the receiver passband be less than the ringing time of the receiver (approximately ½ BW). It should be repeated here that if the radar is unprotected by pulse interference rejecting or similar A-J circuits this requirement probably does not apply against a completely automatic tracking radar.

Against a well protected radar the above requirement may very considerably decrease the effective width of a jamming barrage produced by FM-by-noise. For example if gaussian noise is used as a modulating signal the average rate of passes through a receiver passband falls off fairly rapidly as the receiver is tuned away from the center of the barrage (References 5 and 38). In addition to this the power density spectrum of the barrage will be approximately gaussian shaped so that the jamming power intercepted by the victim receiver also falls off at and is tuned away from the center of the barrage (References 5 and 12). It is a fairly simple matter to change the character of the modulating noise in such a way as to result in a prac-

tically uniform power density across the entire jammer output barrage, but this causes the average rate of passes through the carrier to fall off even faster toward the edges of the barrage (Section 14.2.4). The wide band widths of most tracking radars make it difficult to solve this problem by the most straightforward approach of including such high modulating frequencies that the average rate of passes remains sufficiently high across most of the band. Other solutions are being investigated (Reference 38) such as those suggested by Middleton in Reference 5.

14.3.2.2 Velocity Trackers

In some tracking radar applications, notably missile seekers, precise unambiguous range information is not required and target discrimination may be accomplished on the basis of the doppler shift of the echo. This allows the use of a relatively simple type of radar, the c-w doppler radar, a block diagram of which is shown in Figure 14-14. As the name implies, the trans-

FIGURE 14-14 A simplified c-w doppler radar

mitted signal is unmodulated (except perhaps for coding purposes). The return signal is shifted in frequency by an amount depending on the radial velocity of the target. The speed gate shown within the dotted lines is a servo loop which in effect tracks the doppler shift with a very narrow filter. The angle tracker responds only to those signals which indicate the proper radial velocity.

Because of the 100 percent duty cycle of the transmitter, time sharing between transmitter and receiver is not possible and a serious isolation problem results. The transmitted power of such a radar must be limited to a relatively low level to prevent blocking of the receiver.

To avoid this isolation problem some doppler radars have pulsed transmitters with a receiver blanking gate during the transmitted pulse which allows much higher transmitted power. In order to avoid doppler ambiguities these radars must have a very high pulse repetition frequency and hence a very high duty ratio. For practical purposes doppler tracking is usually done only on the main line of the received signal spectrum.

Range information can be obtained with a pulse doppler radar at the expense of considerable complication. A method of range tracking which is
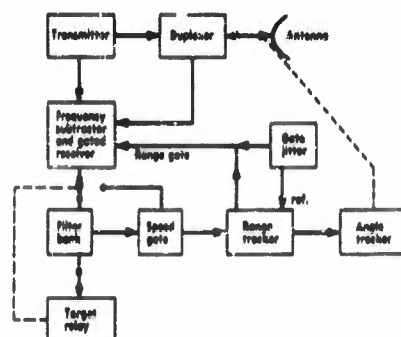
FIGURE 14-15   A simplified pulse doppler radar

sometimes used is shown in Figure 14-15. Since only one line of the signal spectrum is passed by the speed gate, the signal at the range tracker is no longer a pulse. Range error is determined by time jittering the gate applied to the signal in the receiver. This imposes on the signal an amplitude modulation which is retained on the output of the speed gate. This modulation is analyzed in the range tracker to determine the magnitude and direction of the range error, and the appropriate correction made to center the range gate on the target pulse.

Because of the high prf the range information is ambiguous. If unambiguous range is required the ambiguities can be resolved by switching the prf. Another effect of the high prf is eclipsing of the return signal from one transmitted pulse by a later pulse. In some models eclipsing is avoided by varying the prf in such a way that the return signal always falls between two transmissions.

The most important difference between the two types of doppler radar from the ECM viewpoint is the higher power of the pulse type, and jamming techniques designed to be used against them are quite similar.

Because of the extremely narrow bandwidth of the speed gate filter, barrage jamming of either the C-W or pulse doppler radar would be very inefficient. Spot noise jamming also would seem to be very difficult because of the required accuracy in frequency set-on.

Repeater jammers naturally satisfy the accurate frequency set-on requirements and at the same time lend themselves very well to deception techniques. Repeaters are described elsewhere in this volume and so will be discussed very briefly.

Since the doppler radar achieves target selection by tracking the frequency shift of the target echo, one goal of the jammer is to cause the speed gate to move away from the true doppler frequency. This can be accomplished by imposing a false frequency shift on the repeated signal, starting the shift at zero and gradually increasing it. If the frequency shift is then abruptly reduced to zero again the speed gate will no longer be tracking and reacquisition is necessary.

The requirements on the jammer to accomplish this speed-gate grabbing are only that the repeated signal exceed the normal skin return from the

target and that the pull-off rate be kept within the acceleration capability of the victim radar tracking loop.

If the victim radar is a pulse-doppler type (References 39 and 40) so that range information is available (even though it may be ambiguous), range rate may be determined and compared with the apparent radial velocity indicated by the doppler shift. If the tracking loop is made to ignore doppler shifts which do not agree with the range rate measurement the gates would remain locked on the true doppler shift.

One suggested jamming technique for countering this CCM is to use a pulsed repeater and vary the delay of the repeated pulse to cause an apparent change in radial velocity. This apparent rate of change of range is made to agree with the false doppler imposed on the repeated signal. In certain applications the doppler shift of the target echo may be used for other purposes than target discrimination. For example the detonation of a guided missile fuze is often controlled by the indicated radial velocity of the target. In such a case jamming may be more effective if the false doppler signal is programmed to cause premature detonation of the missile warhead.

Since the angle tracking loop of a doppler radar derives error information from only that part of the signal which is passed by the speed gate, capture of the speed gate makes possible the introduction of false angle information also (References 41 and 42). This may be accomplished by modulations discussed in the following sections.

14.3.2.3 Angle Trackers

Conical-Scan Tracking. The effect of jamming signals on conical-scan angle trackers will probably continue to be an important consideration in the design of countermeasures systems to be used against airborne tracking radars, particularly missile seekers, since there appears to be a strong preference for conical-scan tracking in these applications.

The operation of a conical-scan tracker is shown in Figure 14-16. Although target selection is achieved in this example by range gating it may also be done on the basis of radial velocity as described previously.

As the antenna rotates, the narrow pencil beam scans a cone shaped pattern. If the target is not on the axis of the scanned cone, the amplitude of the signal return will vary sinusoidally at the scanning frequency. The magnitude of the variation is proportional to the magnitude of the error, and the phase of the variation, with respect to reference signals generated within the radar, indicates the direction of the error. The error signal and reference signal are used as shown to position the antenna mount in a direction to reduce the error to zero.

If the radar is of the pulse type, the angular error is derived from that
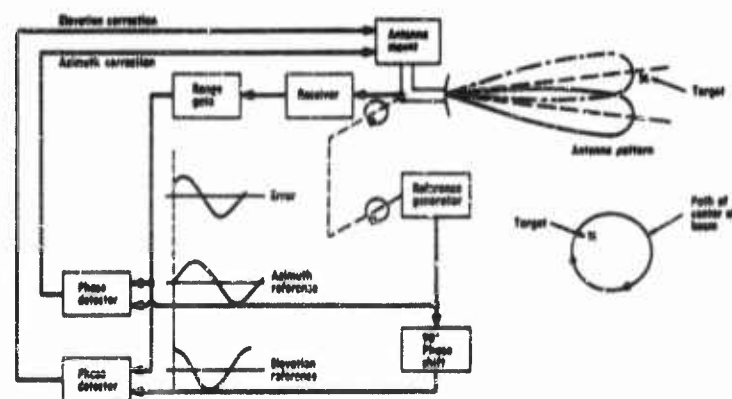
FIGURE 14-16   Conical-scan angle tracker

part of the return signal which has been selected by the range gates. This is necessary so that returns from objects in approximately the same direction, but at different ranges will not influence the tracker.

In order to counter such a radar a jammer must either prevent the radar from detecting the amplitude variations which indicate an error or cause within the radar an apparent error signal which will be acceptable to the error detecting circuits but which differs from the true error in such a way that the tracking error is increased. The latter method will be referred to as deception jamming.

Examples of the first, or masking type, of jamming are c-w and high-frequency noise such as described above for range tracker jamming. In spite of the agc, it is often fairly easy to saturate the radar receiver with either c-w or noise so that the amplitude variations, which indicate an error, are lost.

There are CCM available which prevent saturation and make it possible for a con-scan tracker to track the noise source itself. However, these anti-jam techniques involve a loss of target discrimination by range gates or speed gates so that although the radar could angle track on a single noise source it should be vulnerable to dispersed noise jammers, particularly during the acquisition phase. Such a dispersion can be accomplished by a number of techniques such as jammer carrying decoys, many aircraft in a formation equipped with jammers, illuminating chaff clouds with noise, etc.

Deception jamming (References 42 and 43) in general requires less power than masking since it is not necessary to saturate the radar.

To use the deception type of jamming against a pulse radar it is not enough merely to amplitude modulate a c-w carrier with the con-scan lobing

frequency since the video amplifier of the radar will usually reject such low frequencies. It is necessary to superimpose the false error signal on a modulation which will be passed by the radar video amplifier. This latter modulation can of course be pulse synchronized with the radar pulses as in a repeater or some other type of high frequency modulation such as that used against the range tracker above.

The low frequency false error modulation can then be superimposed by amplitude modulating at the desired frequency or by frequency modulating at half the desired frequency with a deviation of several receiver bandwidths.

In order to implement the deception type of jamming most efficiently, the scanning rate of the radar antenna must be known. If the radar uses the same antenna for transmission and reception, the radar pulses incident on the target will vary in amplitude at the scanning rate whenever there is a tracking error. If, however, the radar is semiactive, that is the transmitter is located remotely from the receiver as in the case of many missile-borne homing radars, then the transmitting antenna is not scanned and it is very difficult, if not impossible to observe the radar scan frequency. Some later models of active conical scan radars have been designed to prevent scanning of the transmitted beam. With sufficient intelligence information it may be possible to predict, within sufficiently narrow limits, the scanning rate of a radar likely to be encountered in a given situation.

For convenience of discussion we will consider three different cases: (1) The frequency and phase of the victim radar scan rate are observed; (2) The scan rate is not observed, but a probable scan rate is known within say a few percent and (3) No information is available regarding the scan rate.

In case (1) the optimum false error signal with which to modulate the jamming is obviously a sine wave at the scan frequency but differing in phase. In case (2) the band of possible scan frequencies can be covered either by slowly sweeping the jammer modulating frequency or by modulating with a narrow band of noise centered at the most likely frequency.

In case (3) it can only be assumed that if the victim radar is a con-scan radar its scan-frequency is somewhere between 20cps and several thousand cps. It is not practical to modulate the jamming signal with a band of noise whose spectrum covers this wide range of frequencies since the radar angle tracking servo responds only to modulation at frequencies within about one cycle per second of the scan frequency, so only a very small percentage of the jamming power would be effective. Similarly, it would not be practical to sweep the modulating frequency over the band since the response of the servo is such that times of the order of one second are required to cause a substantial error. If the sweep speed is slow enough so that the modulating

frequency is within the passband of the servo for about one second, it would take about 15 minutes to cover the range of possible frequencies, and even then there would be an error caused for only about two seconds of each sweep. It appears that in this last case the deceptive type of angle jamming has little value although the masking type mentioned previously would still be effective, causing the radar to lose range information and to sacrifice target discrimination in order to home on the jamming signal. Since in this case the signal amplitude variation which is characteristic of con-scan radars is not observed, one would probably not know whether the victim radar is a conical-scan or a monopulse radar and the best approach would seem to be to treat it as if it were the latter. Countermeasures against monopulse radars are discussed in the following section.

Monopulse. The sequential nature of error detection in the conical scan tracker sometimes results in noisy tracking and is also the source of a vulnerability to certain types of jamming. In a monopulse tracking radar, angle information is derived from each pulse. Monopulse radars fall into three groups—amplitude comparison or simultaneous lobing, phase comparison or interferometer, and a combination of the first two.

An amplitude comparison monopulse radar is shown in Figure 14-17. Two identical channels are required to track in both azimuth and elevation. Each channel has two antennas, cocked slightly so that their center lines form an angle as shown. An angle error is indicated by a difference in amplitude of the signal return in the two channels. The amplitude comparison can be made at any stage of the radar (r-f, i-f, or video). In the example shown the comparison is made at r-f, while the actual error is determined by phase detecting the difference signal with the sum signal.

In a phase comparison monopulse, Figure 14-18, the two horns of a channel are placed so that their centerlines are in the same direction but the horns themselves are displaced by several wavelengths. If a target is on the centerline the return signals to the two horns will be in phase. If the target is off the centerline the signal will reach one horn before the other and the horn signals will be out of phase by an amount depending on the angular error and the horn separation.

The above systems require four antennas to track in azimuth and elevation. A system has been designed which reduces the number of antennas to two by combining the amplitude and phase comparison types. If two antennas are mounted so that they are displaced in the horizontal (Figure 14-19) and their center lines form an angle in the vertical plane, then azimuth elevation error will be indicated by a phase difference and elevation error by an amplitude difference.
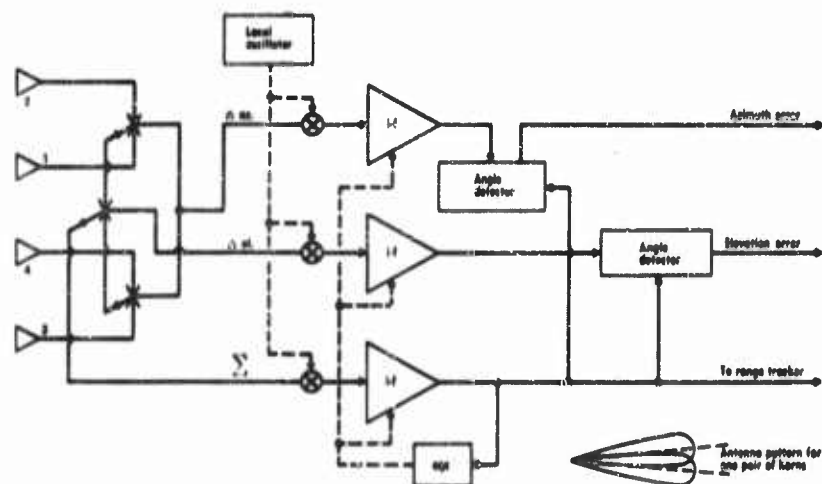
FIGURE 14-17   Amplitude comparison monopulse

Of the three types described, amplitude comparison appears to be by far the most common. The complexity of the waveguide circuitry of monopulse radars has discouraged their use in some airborne applications, particularly missile seekers, but they are widely used in ground based fire control systems.

Jamming techniques against monopulse radars can be divided into the following groups (Reference 44):

A. Off-Frequency Jamming (which depends on imperfections of design or construction)
     1. Image Jamming
     2. Skirt Frequency Jamming
B. Multiple Source Methods
     1. Blinking
     2. Formation Jamming
     3. Cross-Eye

Image jamming (Reference 45), that is jamming at the image frequency of the radar, depends on the fact that the phase angle at i-f between two signals at the image frequency is the reverse of that which would appear at i-f if the two signals were at the normal response frequency of the receiver. Since the phase comparison monopulse determines the direction of the error by the direction of the phase error between the two horn signals, image jamming causes the antenna to be driven away from the target if the jamming power exceeds signal power.

FIGURE 14-18   One channel of a phase comparison monopulse.

FIGURE 14-19   Combination monopulse
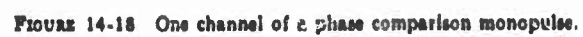
In some monopulse radars slight differences in length of signal paths from the horns are compensated for by phase shifts at i-f. For signals at the image frequency however, the compensating phase shift adds to the original phase shift and the original error is doubled instead of being canceled. This effect can be obtained even in amplitude comparison radars if the compensated error is large enough.

Image rejection in the radar would eliminate, or at least make more difficult, effective image jamming.

Skirt frequency jamming refers to jamming at frequencies on the skirts of the response curve of the radar. Its effectiveness depends on unbalances between sum and difference channels at these frequencies where rapid phase shifts are present in each channel. Of course it can be effectively countered by careful design and construction of the radar.

A well designed monopulse radar can track very effectively on a single source of noise jamming (Reference 46). However, in most applications of monopulse, range information is also required and as discussed previously, noise jamming is an effective countermeasure against range trackers.

Whether target range or doppler shift is used for target discrimination, noise jammers located remotely from the target will affect the angle tracking circuits and cause an error since neither time nor frequency gating will favor noise radiated from the target. Dispersed jammer techniques should be as effective against monopulse as against conical-scan radars which were discussed previously.

Blinking refers to jamming (either noise or repeated signal) alternately from displaced points in an aircraft or from separate aircraft. The effect is to increase the wander of the tracking point.

Cross-eye (References 47, 48, 49, 50, and 51) or phase front distortion is a multiple source technique in which radar signals received at one point in an aircraft are amplified and retransmitted from a second point as far as possible from the first, while signals received at the second point are shifted in phase by 180 degrees and retransmitted from the first point. Analysis of the technique is quite complicated but it appears that a large jamming to signal ratio is required to be effective. Because of inherent delays in the retransmission, the technique can be countered by obtaining angle information from only the leading portion of the signal pulse.

A summary of the above techniques is given in Figure 14-20, which was taken from a paper (Reference 44) by Walters of the Johns Hopkins University Radiation Laboratory.

Type of Jamming	Type of Monopulse A = Angle P = Phase C = Combination	Effect below Threshold	Power J/S Threshold	Effect above Threshold	Countercountermeasures and Comments
Image	P C	Noise tracking (small due to limit by servo bandwidth)	1	Error reversal	Image rejection (by rf preselection dual-channel IF or high IF)
	A		1 and must have 45° rf error compensated or IF		Absorb or reduce rf channel length difference to below 45°
Skirt	A P	Loop gain degraded	Large	Error reversal	More nearly equal sum and difference channels
	C	and crosstalk increased			
Noise	A P C	None	See discussion		Ideal monopulse will track noise
Blinking	A P C	None	1	Enhanced tracking point wander	Applicable to missileborne seekers
Crosseye	A P C	Noise tracking	Approx. 10	Error reversal	Pulse edge angle determination (to take advantage of repeater delay time)

FIGURE 14-20 Summary of monopulse jamming techniques

14.3.3 Experimental Methods of Determining Jamming Effectiveness

Simulation. Because the weight or space allocated to ECM equipment is necessarily limited, and since the weight of the equipment increases rapidly with an increase in power output capability, it is important that the jamming signal radiated by the equipment be designed for maximum effectiveness, for a given amount of power. Although field testing is essential as a final step in verifying the effectiveness of jamming equipment, reliable data regarding the relative effectiveness of various types of jamming signals is best obtained in a laboratory type test (References 1, 12, 31, 36, and 37). Since the parameter of principal interest is the ratio of jamming power to signal power required to produce a given degree of jamming of a particular type of radar, it is desirable to control as many as possible of the other variables involved such as: target velocity, range and aspect; jammer antenna pattern; signal and jamming propagation; ground clutter, etc.

In the field it is very difficult to reproduce test conditions several times in order to determine the relative effectiveness of several types of jamming signals, whereas in the laboratory it is relatively simple to hold constant all of the parameters which would affect the comparison.

While the ultimate in versatility and control of conditions is probably achieved by complete simulation of the radar by means of an analog computer, it often requires a sacrifice in realism, for example, the antenna and r-f circuitry of the radar and jammer may be over-simplified.

A good compromise between realism and versatility appears to be to set up an actual radar on a test range with a reflector or a repeater to provide a target signal (References 31 and 52). The jamming signals to be evaluated can be transmitted to the radar from the same location as the simulated target to provide a completely closed loop test. It is a relatively simple matter to instrument the radar and signal source to indicate the actual error produced by jamming as well as the tracking error present without jamming.

Test Range. An example of such a test range is the one at the Johns Hopkins University Radiation Laboratory. The range is indoors and about 40 feet long which is sufficient to avoid near field antenna effects with small airborne radars. In order to simulate targets at realistic ranges an AN/UPM-11 range calibrator is used. This device mixes the signal from the radar with a local oscillator signal, delays the i-f signal with a quartz delay line and uses the same local oscillator to convert the delayed i-f signal back to the original frequency. With this device any modulation of the illuminating energy (due for example, to scanning of the radar antenna) is retained in the simulated echo signal.

If a realistically noisy target signal is desired the AN/UPM-11 has provisions for amplitude modulating the repeated signal in accordance with an external voltage.

Since the delay line in the UPM-11 is fixed, target range motion is simulated by delaying the synchronizing pulse in the radar and varying the delay time. Angle motion of the target is accomplished by attaching the horn of the AN/UPM-11 to the end of a movable arm and feeding the horn through a rotating feed.

Because of the short ranges involved the radar transmitted power is reduced by means of a directional coupler and a dummy load. The short range also reduces the required power level of the jamming sources, and makes possible the testing of some jamming signals which may not yet be available at power levels required under field conditions.

For example using a low power klystron as the primary source it is relatively easy to produce (References 31 and 30): frequency modulation of the

klystron by a wide variety of waveforms, including noise, up through video frequencies; amplitude modulation at high audio frequencies by using ferrite modulators; suppressed carrier AM by video frequency or i-f signals or noise by means of balanced mixers; single sideband, suppressed carrier AM by video or i-f signals or noise by means of a modified balanced mixer.

Of course other jamming signals can be generated using traveling wave tubes and carcinotrons, etc.

Measurement of Jamming to Signal Ratio. The ratio of total jamming power to peak signal power is easily measured at the feed of the repeater jamming horn. If a calibrated attenuator is used to vary the jamming level, the jamming power can be read with a bolometer at one attenuator setting and other settings related to it.

One technique for measuring signal power at this point is to fix the radar receiver gain, noting the peak signal amplitude at video, then substituting a calibrated signal generator for the UPM-11 and noting the output level at which the signal produced at the radar video equals that previously obtained from the UPM-11. The signal generator can be easily checked against the bolometer used for the jammer power measurement.

If a comparison is desired of signal power and that portion of the jamming power which falls within the radar receiver passband, the mean square output of the radar i-f amplifier can be measured with a cathode follower and an r-f thermocouple. Since only a comparison of signal and jamming power is required, it is not necessary to calibrate the device. With the jammer turned off a signal power reference is obtained by noting the thermocouple output due to a c-w signal of the same level as the peak of the pulse signal used above. The jammer output level is adjusted with the calibrated attenuator to cause the same thermocouple output (with the signal off) and the attenuator reading noted. With this attenuator reading giving unity J/S, other readings can easily be related to it. Of course saturation of the radar receiver should be avoided during the power measurement.

14.4 The Communications Jamming Problem

The problems of communications jamming differ from those of radar jamming in several respects. For example, the jamming transmitter does not necessarily have a distance advantage over the desired signal transmitter in the communications case, whereas in the radar case the jamming transmitter can take advantage of the two-way transmission requirements. Further, the communications signals in general have had their power concentrated in much smaller portions of the spectrum, although with spread-spectrum techniques seeing increasing application, this may not be true in the future.

Two modes of radio communications which are important in modern warfare are voice and coded data links. In the gross aspect, both of these modulations are of continuous-wave type rather than being of a pulsed character.

14.4.1 Data Link Jamming

It is difficult to generalize about the characteristics of pulse-coded data link systems. Less laboratory evaluation work has been done with ECM than in the voice communications case. It can probably be assumed that the designers of these systems have taken precautions to use utmost time and bandwidth economy to transmit the desired message. However, they still have the advantage in wide latitude of trade-off between message speed and bandwidth. Military data-link systems in the United States have shown individual channel bandwidth differences of a thousand to one, with corresponding variations in total message time. Systems in which individual channel bandwidth is very small generally employ frequency multiplexing as well as time multiplexing to reduce the complexity of the system whereas with broadband systems, time multiplexing suffices.

Present systems are almost completely nonredundant because of the elementary state of the data-link art, but future systems may well contain as much redundancy as is consistent with the amount of traffic to be handled by each channel. Presently, since the systems are able to accept only signals conforming to a rigid set of time specifications, they are vulnerable to any sufficiently strong extra energy pulse placed in the message bandwidth during the message time. The effect of such interference is to cause rejection of all signals so disturbed. It is very doubtful whether "intelligent" jammers which falsify the victim message are feasible. This is because of the high degree of intelligence information which would be required and because of the high degree of timing accuracy with which the interfering signal would have to be injected in order not to interfere with the various parity checks in the victim signal.

The addition of higher power transmitters and more directive antennas to the data-link systems will result in proportionately higher power required of the ECM oment, particularly in situations where receipt of an answer by a ground station is not necessary for continuous information flow. Ground station powers may eventually become so great that airborne data-link jammers are no longer feasible.

14.4.2 Voice Communications Jamming

The problem of denying the enemy utilization of the electromagnetic spectrum for communication purposes is, like other aspects of the countermeas-

ures-countercountermeasures game, changing as new communications techniques become an integral part of the military picture. The development of spread-spectrum and noiselike modulation schemes, together with more sophisticated correlation methods for collapsing a broad spectrum back down into a narrow bandwidth and bringing the signal up out of the natural noise or the jamming signal have changed the jammer's problem from that of intercepting and identifying a fairly narrow-band, high signal-to-noise ratio situation to one of distinguishing a pseudorandom process from truly random noise and devising a jamming technique which will be effective in rendering such a signal useless at the intended receiver where an appreciable amount of *a priori* information regarding the signal exists. Such trends require that the jammer have available more complete information on the target signal in order to accomplish its mission. Sophistication in jamming techniques becomes necessary simply because of the difficulty of utilizing brute force methods throughout the frequency spectrum.

However, considering the very large number of communication systems employing conventional AM and FM methods, it is well to examine also the jamming of such systems and the attendant problems of evaluating jamming effectiveness.

14.4.2.1 Theoretical Considerations

Hok (Reference 53) has shown that the theoretically attainable jamming effectiveness grows directly with the available statistical and exact historical information about the target signal. Utilization of this information in production of an effective jamming signal results in specialized and elaborate jamming equipment with a resultant increase in the equipment's vulnerability to changes introduced in the target system. Repeaters and memory devices represent a trend toward such equipments insofar as they attempt to make use of the information available to them and try to become partially predictive in their operation.

The study of optimum jamming signals is dominated by those signals having a gaussian amplitude distribution because it has the highest entropy of any distribution of given average power, as well as providing the largest entropy intersection attainable with any other energy distribution, such as that of the target signal (Reference 53). The effect of the jamming is, desirably, to maximize the uncertainty at the receiver as to what target signal was transmitted.

The generation of gaussian noise at appreciable power levels in various portions of the communication spectrum has always been a major problem and so studies of the effectiveness of other jamming signals, such as are represented by carrier waves modulated by video frequency gaussian noise or

other signals have been of interest (References 54, 55 and 56). Two of these signals are commonly referred to as AM-by-noise and FM-by-noise.

On a theoretical basis, Stewart (Reference 56) has concluded that FM-by-noise is superior to AM-by-noise for general-purpose jamming of either AM or FM communications, particularly from the standpoint of existing techniques of power generation.

Theoretical studies of the effectiveness of jamming signals may logically be expected to have made excellent use of the work of Rice and Shannon (References 57 and 58). Certainly the reduction in the capacity of a channel to transmit information is a suitable measure of the effectiveness of jamming or of the resistance of the channel to jamming activity. This point of view was manifested in part in Chapter 8 by Tanner in the discussion of "A System Approach to the Countermeasures Problem." The experimental justification of the methods is not quite so clear-cut because of the difficulty in controlling or specifying all of the variables involved in test runs employing willful operators. Approaches to this problem are treated in the next section.

14.4.2.2 Evaluation of Jamming Effectiveness

Each of the several constraints characterizing military communication systems—signal-to-noise ratio, limited message ensemble size, coding operations, feedback from receiver to transmitter, redundancy, time for message transmission, synchronization, type of modulation, directivity of the communication link, etc.—has a bearing on the jamming effort required to be effective. A simple "yes" or "no" is more difficult to jam than a quotation from a field manual. If the receiver can request a repetition or if the transmitter can switch to Morse code on demand, the jammer's task is changed. It therefore becomes necessary to define a meaningful test which will yield valid as well as reliable results.

The time-honored method of checking communication systems is the use of the articulation test. Extensive laboratory testing programs (References 59, 60, 61, 62, 63, 64, and 65) have been carried out using the conventional test procedure in which a receiving operator writes down each isolated word he hears over the system under test and a percentage correct is obtained as the measure of transmission success in the face of jamming or interference. General Electronic Laboratories has developed an articulation testing device which yields an over-all evaluation of the system without the necessity for either a transmitting operator or receiving observer (References 59 and 63).

One approach to the problem attempts to remove some of the objections to the articulation test procedure, these being the lack of consideration for the time of message transmission, the failure to allow structure or redundancy in the transmitted message, and the lack of opportunity for feedback to

be incorporated. All of these characteristics appear in military transmissions. A consequence of these considerations was the development of the "Map Test" at the Electronic Defense Group at the University of Michigan. The test is adequately covered in References 66, 67, 68, and 69. The basic idea is represented in Figures 14-21 and 14-22. The transmitting observer selects
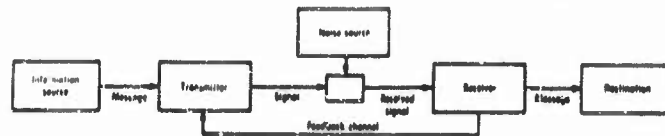
FIGURE 14-21 Basic communications system
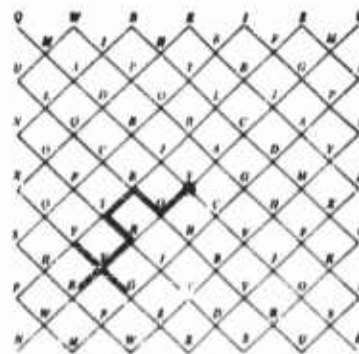
FIGURE 14-22 The EDG diamond map 4 for six town routes

one of a limited ensemble of 972 possible routes on a grid of "towns" whose names are the letters of the phonetic alphabet. Each route of six towns represents a fixed amount of information, 9.925 bits, to be transmitted in a minimum (and measured) amount of time when jamming is being utilized. The receiving operator records the route received on a similar grid. Various forms of feedback may be employed, such as a noise-free channel, a jammed channel, a low-capacity channel with a light to signal the transmitting operator to proceed with the transmission. The limited choice available to the transmitter at each grid intersection is representative of the structured redundancy of the map; the receiving operator knows that only a limited choice is possible. The measured time for transmission can be taken as the jamming effectiveness measure for each J/S ratio and for each experimental arrangement of parameters. By way of reference, a six-town map can be transmitted in the clear in about two seconds by trained observers; if they require more than twenty seconds for transmission under jamming conditions, they stand a very small chance of getting the message through at all. The

test gives excellent quantitative results in the laboratory. Some examples of the effectiveness of various types of jamming of AM and FM systems as found in this type of test are given in Figures 14-23 and 14-24.
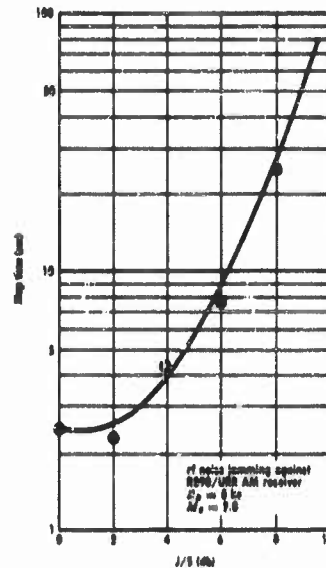
FIGURE 14-23   Standard map test   FIGURE 14-24   Pulsed FM-by-noise jamming
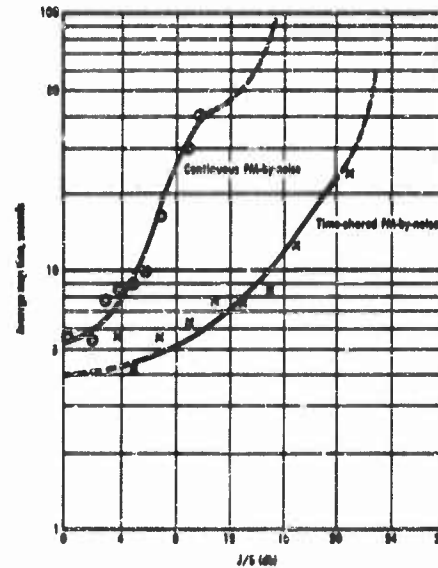
The jamming of conventional modulations such as AM and FM has been well covered since World War II (References 70, 71, 72, 73, 74, and 75) and jammers employing noise or other random-character modulations for frequency modulation of an r-f carrier have been shown to be effective under ordinary circumstances. The effects of lack of set-on accuracy on the part of the jammer (Figure 14-25), the determination of optimum jammer parameters for specified situations, and the effects of receiver counter-countermeasures tactics have been studied in detail. While no one set of numbers can be employed to cope with all field situations, the Figures 14-25 and 14-26 represent some special cases, the former illustrating the effect of jammer offset in frequency with modulation parameters optimized for each misalignment value when working against an AM receiver, the latter showing a case against an FM receiver, the capture effect being noticeable for the continuous jamming signal but being comparatively ineffective when used in a 50 percent duty cycle. It may be expected that actual field situations will require lower values of J/S because of the many nuisance factors at work against the target channel.
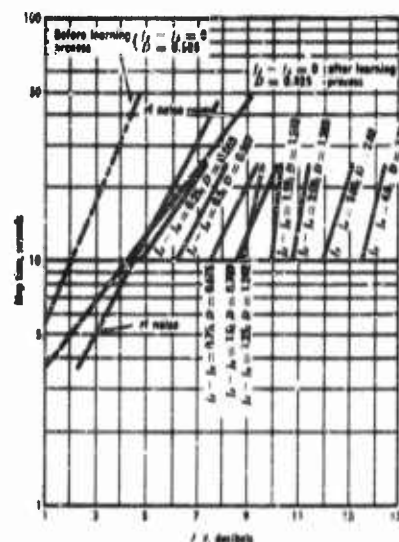
FIGURE 14-25 FM-by-noise jamming

FIGURE 14-26 Map test results comparing continuous and interrupted FM-by-noise jamming against FM receiver

Of concern in recent years has been the discovery that a target FM receiver may detune from his own intended transmission, and thereby from the jammer who can only assume a properly tuned system against which to operate, and be able to read through extremely high jamming levels (in terms of J/S at the receiver terminals) with little difficulty. This situation is depicted in Figure 14-27 which shows laboratory data for a working r-f link subjected to

FIGURE 14-27 Map test results showing the possible severity of the detuning effect

FM-by-noise jamming. It is apparent that the effectiveness of the jamming decreases as the receiver is detuned to the region of the back slopes of the over-all receiver characteristic. In view of the very large amount of FM equipment in use, this effect is a significant threat. Hellwarth, Rothschild and Ross have studied this situation and concluded that a time-sampling phenomenon is at work, permitting samples of the intended transmission to be taken at times when the

jamming signal is out of the passband of the receiver. Statistically, this

occurs often enough for normal modulation schemes so that the receiver is able to construct an intelligible message at large J/S ratios. This work is treated extensively in References 76 and 77.

A means of combating the detuning effect, at the expense of jammer complexity and J/S ratio, is to arrange to have a jamming signal in the receiver's passb' id all of the time (Reference 76). Figure 14-28 illustrates
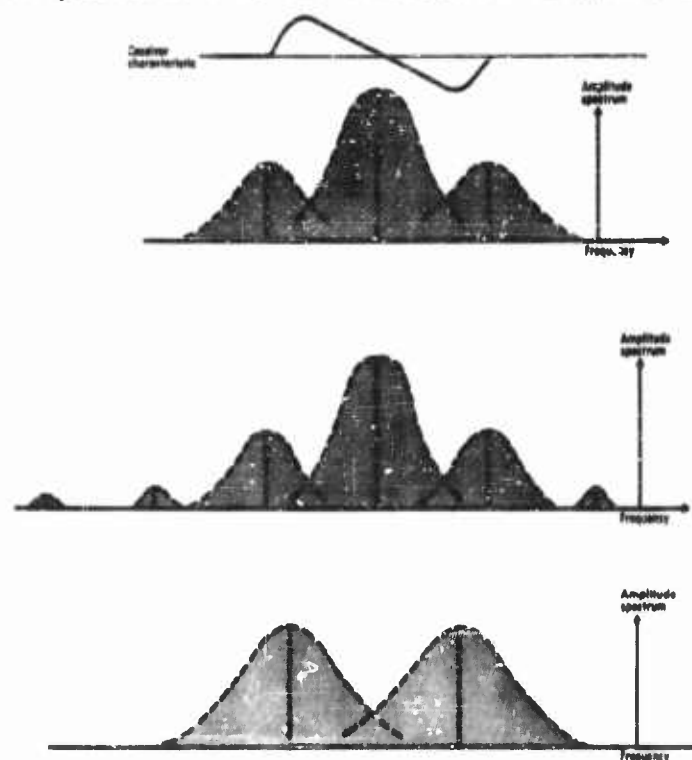
FIGURE 14-28   Amplitude spectra of jamming signals used to defeat detuning

three techniques for accomplishing this task. While the detuning scheme cannot be employed against r-f noise jamming, or DINA, a barrage jammer of this type requires several times as much power as the first two alternatives given in the figure. The third approach requires two separate transmitters, synchronously modulated by a video noise signal.

Detuning of an AM receiver is not an effective counter-countermeasures technique.

Because look-through on the part of the jammer is often desired, and

because all look-through equipments involve a sampling of the target transmission, the question of effectiveness of jamming which is time-shared arises. It has been shown (Reference 78) that the effect of one type of time-sharing of FM-by-noise is to degrade the effectiveness of the jamming. Figure 14-29 illustrates this situation for various duty cycles of the jammer.
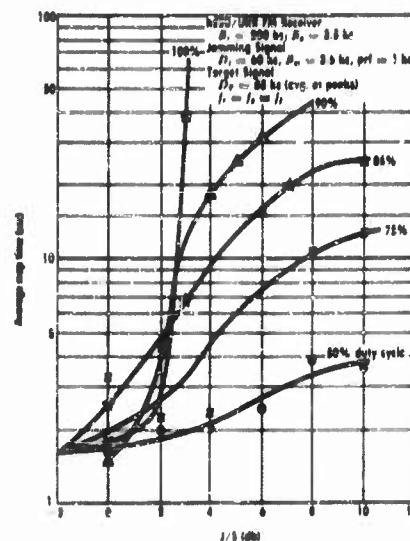
FIGURE 14-29 Time-shared FM-by-noise jamming effect of changing duty cycle

The desirability of having some knowledge of target-link behavior when subjected to jamming led to the development of a continuous-monitor jamming system which used the jamming signal as a local oscillator and synchronous control signal so that a visual display provided continuous information as to whether the victim had left the air or not. Another feature was audible read-through so that an additional evaluation of the effectiveness of the jamming could be ascertained from the victim transmission and feedback. This work is covered in Reference 79.

At this writing two of the current problems in the study of jamming effectiveness center around single-sideband and pseudorandom transmissions. The former method of radio communication is well-known, although the problem of jamming such signals is only now becoming of concern because of the increasing adoption of SSB techniques by military organizations.

Pseudorandom techniques are being found increasingly in radar, communications and control systems (Reference 80). They are being used for one or more of three possible special objectives: an A-J advantage, signal hiding, or message security. The first objective concerns working satisfactorily in the presence of a large amount of jamming. The second considers an attempt to avoid detection by a nonintended receiver. The third relates to encoding properties which may accrue from the use of deterministic signals resulting from an application of a linear shift-register-generated sequence.

In one form of employment (Reference 81), "either one of two waveforms is broadcast from the transmitter; the receiver, which is generating the exact same sequence in synchronism with the transmitter, crosscorrelates its inter-

nally generatea waveform with the incoming signal to determine which of the two was broadcast. The peculiar advantages of the digitally-generated waveform in this application are that the resulting waveform has a wide bandwidth, the power is nearly uniform throughout the bandwidths, and the autocorrelation function is approximately zero outside of the narrow peak at zero delay."

It is apparent that the intended receiver already has a great deal of knowledge about the signal being transmitted. This aggravates the jammer's problem in many cases if synchronism has been established because of the possibility that the signal may be spread over an appreciable portion of th spectrum or may be so far down in the ambient noise as to be difficult to detec and identify as one which should be jammed. Such a sophisticated system wh.   uire sophisticated jamming techniques if it is to be other than brute force in ..

The bibliography on this chapter will be found to contain many references whose titles will suggest other aspects of the jamming effectiveness studies which have taken place in recent years.

REFERENCES

1. Lawson, J. L. and G. E. Uhlenbeck, "Threshold Signals," McGraw-Hill Book Company, Inc., Massachusetts Institute of Technology, Radiation Laboratory Series, Vol. 24, 1950.

2. Marcum, J. I., "A Statistical Theory of Target Detection by Pulsed Radar," Project RAND, Douglas Aircraft Company, Inc., RA-15061, December 1945. (CONFIDENTIAL)

3. Hok, G., "A study of Optimum Jamming Signals," The University of Michigan, Electronic Defense Group, Tech. Report No. 17, October 1954. (SECRET)

4. Middleton, D. and D. Van Meter, Detection and Extraction of Signals in Noise from the Point of View of Statistical Decision Theory, *Society for Industrial and Applied Mathematics*, Parts I and II, 3 December 1955 and 4 June 1956.

5. Middleton, D., "On the Problem of Jamming Criteria in Radar ECM with an Application to Barrage Jamming," The Johns Hopkins University, Radiation Laboratory, Tech. Report No. 34, September 1956. (CONFIDENTIAL)

6. Baker, C. H. and G. B. Thornton, "A Guide to Factors Affecting Radar Operator's Efficiency," Defense Research Board, Department of National Defense, Canada, December 1953. (SECRET)

7. Williams, S. B., "Visibility on Radar Scopes," Human Factors in Undersea Warfare, National Research Council, Washington, D.C., 1949.

8. Morgan, C. T., "Theory and Problems of Radar Visibility," Naval Research Laboratory Report No. 3965. April 1952.

9. Blackwell, H. R., "Psychophysical Thresholds—Experimental Studies of Measurements," University of Michigan Press, Ann Arbor, 1953.

10. Derr, V. E., "Interim Report on the Experimental Search Radar Jamming Program," The Johns Hopkins University, Radiation Laboratory, Tech. Report No. 19, May 1955. (SECRET)

11. Wild, T. A. and V. E. Derr, "The Effectiveness of a Carrier Frequency Modulated by Noise in Jamming the AN/FPS-3 Receiver," The Johns Hopkins University, Radiation Laboratory, Tech. Report No. 35, September 1956. (SECRET)

12. Wild, T. A., and L. K. Lauderdale, "Final Report on the Experimental Search Radar Jamming Program," The Johns Hopkins University, Radiation Laboratory, Tech. Report No. 45, February 1958. (SECRET)

13. Wild, T. A. and A. B. Novikoff, "Comparison Tests Between RL Observers and AF Radar Operators," The Johns Hopkins University, Radiation Laboratory, IMA-32, December 1955. (SECRET)

14. Rothschild, D. R. et al, "Study of Jamming Effectiveness and Vulnerability," The University of Michigan, Electronic Defense Group, Progress Report No. 6, Task EDG-7, December 1956. (SECRET)

15. Brown, W. M. "Noise Statistics after Transformations Commonly found in Circuits," Tech. Report No. AF-22, OTS Identification Number PB 121515, U.S. Department of Commerce, Washington, D.C.

16. Davenport, W. B. and N. L. Root, "Random Signals and Noise," McGraw-Hill Book Company, Inc., 1958.

17. Wild, T. A., "A Practical ERFER Device," The Johns Hopkins University, Radiation Laboratory, IMA-19, November 1956. (CONFIDENTIAL)

18. Silfer, C. F. "Glossary of Anti-Jamming Techniques and Terminology," Rome Air Development Center, June 1958. (SECRET)

19. "Anti-Jam Techniques Study—Final Report," Contract No. AF 30(602) 1593, Sylvania Electric Products, Applied Research Laboratory, October 1957. (SECRET)

20. "Rome Air Development Center ECCM Symposium"—Four volumes, RADC TR-57-162 A, B, C, and D, ASTIA Document numbers: AD-131308, 09, 10, 11. (SECRET)

21. Karr, P. R., "Possible ECCCM Techniques Against a Class of Jammer Locator Systems," Ramo-Wooldridge Corp., July 1958. (SECRET)

22. Berstein, R. I., "The ORDIR Radar System," Columbia University Engineering Center, Tech. Report T-3/124, May 1958. (SECRET)

23. Harmon, W. W., "The Feasibility of Matched-Filter Radar," Stanford University, Electronics Research Laboratory, Tech. Report No. 35, November 1954. (CONFIDENTIAL)

24. "Research on Anti-Jamming Techniques—Final Report," Contract No. AF33(600)-29326, General Electric Company, Advanced Electronics Center, October 1957. (SECRET)

25. Young, R. D., "Effectiveness of Countermeasures Against Air Defense Systems," Ramo-Wooldridge Corp., November 1954. (SECRET)

26. "Research on Anti-Jamming Techniques," General Electric Company, Advanced Electronics Center at Cornell University, March 1958. (SECRET)

27. "Research on Anti-Jamming Techniques," General Electric Company, Advanced Electronics Center at Cornell University, Interim Engineering Report No. 2, October 1958. (SECRET)

28. Raymond, R. C., "Protection of Radar Target Seekers from the Cyclic Range Gate Stealer," The Rand Corp., RM 1122, August 1953. (SECRET)

29. "Research on Anti-Jamming Techniques," General Electric Company, Advanced Electronics Center at Cornell University, Summary Report, September 1956. (SECRET)

30. Brown, W. M., "On Pulse Train Jamming of Range Tracking Radar," The Johns Hopkins University, Radiation Laboratory, RL/57/IMA-6, August 1957, Research Progress Report Part 1, October 1957. (SECRET)

31. Brown, W. M. and R. H. Richard, "Noise Jamming of Range Tracking Radar Systems," The Johns Hopkins University, Radiation Laboratory, Tech. Report AF-44, January 1958. (SECRET)

32. Ruina, J., "The Effects of Noise on Range Tracking Systems," Microwave Research Institute, Polytechnic Institute of Brooklyn, June 1951.

33. Swerling, P., "Some Factors Affecting the Performance of a Tracking Radar," The Rand Corp. RM 989-1, September 1954.

34. Muchmore, R. D., "Calculation of Range of Automatic Tracking Radars," Hughes Aircraft Co., Tech. Mem. 193.

35. Stone, A. M., "Synthetic Radar Echoes in the Presence of FM Jamming," Massachusetts Institute of Technology, Radiation Laboratory, Report No. 1035, April 1946.

36. Taylor, D. W., "The Effectiveness of AM, FM, and DINA Jamming of the Type A Presentation as a Functional of IF and of Video Overloading," Harvard University, Radio Research Laboratory, Report No. 411-71, January 1944.

37. Taylor, D. W., "Laboratory Studies of Jamming Effectiveness," Harvard University, Report No. 411-248, October 1945.

38. "Research Progress Report, Electronic and Infrared Countermeasures," The Johns Hopkins University, Radiation Laboratory, Report No. AF3-11, October 1958. (SECRET)

39. "Anti-Jamming Design Parameters for Pulse Doppler Radar Systems," Radio Corporation of America, West Coast Electronics Products Dept., July 1958. (SECRET)

40. "Countermeasures Tests of Boeing Experimental Pulse Doppler Radar," Boeing Airplane Company, Pilotless Aircraft Division, Report D51101, August 1956. (SECRET)

41. Wright, M. and W. E. Ayer, "A CW Deception Jammer," Stanford University, Stanford Electronics Laboratories, Tech. Report No. 441-4, December 1957. (SECRET)

42. "Research on FM/CW Repeater," Raytheon Manufacturing Co., Maynard Laboratory, March 1957. (SECRET)

43. Crumly, C. B., "A Preliminary Analysis of Inverse Gain Radar Jamming," Stanford University, Electronics Research Laboratory, Tech. Report No. 10, January 1954. (SECRET)

44. Walters, J. L., "Modulations For Angle Jamming of Monopulse Radar," University of Michigan, Department of Defense Electronic Warfare Symposium, October 1958. (SECRET)

45. Lewis, B. L., "Image Jammers," Naval Research Laboratory, Report No. 5004, September 1957. (SECRET)

46. Barton, D., "Passive Radar Tracking on a Noise Source," RADC ECCM Symposium Transactions, Vol. I, October 1957. (SECRET)

47. Jenks, J. C., "Countermeasures Against Simultaneous Lobing Radars," The Johns Hopkins University, Radiation Laboratory, RL/58/IMA-1, March 1958, Research Progress Report No. AF5-10, April 1958. (SECRET)

48. Walters, J. L., "Angle Jamming of Monopulse Radar," The Johns Hopkins University, Radiation Laboratory, Tech. report in preparation. (SECRET)

49. Lewis, B. L., "An Unusually Effective Radar Countermeasure," Naval Research Laboratory, Report No. 4661, May 1956. (SECRET)

50. Howard, Dean D., "A Further Study of the Cross-Eye Countermeasure System," Naval Research Laboratory, Report No. 4781, July 1956. (SECRET)

51. Thomas, R. K., "Jamming of Simultaneous Lobing Systems," The Johns Hopkins University. Radiation Laboratory, Progress Report No. AF5-3, Part I, October 1956. (SECRET)

52. "Research on Anti-Jamming Techniques," General Electric Company, Advanced Electronics Center at Cornell University, Summary Report, August 1955. (SECRET)

53. Hok, G., "A Study of Optimum Jamming Signals," The University of Michigan, Electronic Defense Group, Tech. Report No 17 (2262-48-T), October 1954.

54. Stewart, J. L., "The Power Spectra of Signals Phase and Frequency Modulated by Gaussian Noise," The University of Michigan, Electronic Defense Group, Tech. Report No. 23, November 1953.

55. Stewart, J. L., "Jamming and Antijamming in Speech Communications," The University of Michigan, Electronic Defense Group, Tech. Report No. 60 (2262-103-T), February 1956.

56. Stewart, J. L., "A Theoretical Evaluation of Frequency-Modulated Jamming Signals Applied Against AM Receivers," The University of Michigan, Electronic Defense Group. Tech. Report No. 25 (1970-6-T), February 1954.

57. Rice, S. O., Mathematical Analysis of Random Noise, Bell System Tech. J., Vols. 23 and 24, 1944 and 1945.

58. Shannon, C. E., The Mathematical Theory of Communication, Bell System Tech. J., July and October 1948.

59. Signal Corps Contract DA-36-039 SC 64460, General Electronic Laboratories, Inc., Cambridge, Massachusetts.

60. Air Force Contract AF 33(616)-115, Cook Research Laboratories, Skokie, Illinois.

61. Subcontract 68-8 of AF 33(616)-68 and 3374-1 of AF 33(616)-3374, Cook Research Laboratories, Skokie, Illinois, under The Johns Hopkins University, Radiation Laboratory, Baltimore, Maryland.

62. Licklider, J. C. R., Articulation Tests, Transactions of 1956 EW Symposium.

63. Billig, L. S., Replacing the Listener in Articulation Testing. *Transactions of 1956 EW Symposium.*

64. Egan, J. P., Criterion in Speech Communication in Noise, *Transactions of 1956 EW Symposium.*

65. Benninghof, R. H., Experimental Studies of Voice Communication Jamming Effectiveness, *Transactions of 1956 EW Symposium.*

66. Birdsall, T. G., The Map Test, *Transactions of 1956 EW Symposium*

67. Rothschild, D. R., Relationship Between Time and Error in Jamming Tests, *Transactions of 1956 EW Symposium*, also see Reference 73.

68. Rothschild, D. R., "Laboratory Methods for the Evaluation of Jamming Signals Against Voice Communication Receivers," The University of Michigan, Electronic Defense Group, Tech. Report No. 59 (2262-120-T), May 1956.

69. Tanner, W. P., Jr., "A Method for Evaluating the Potential of a Signal to Jam Voice Communication," The University of Michigan, Electronic Defense Group, Tech. Report No. 36 (1970-16-T), April 1954.

70. Bacon, F. R. and W. J. Parker, "Investigation of Jamming Effectiveness of FM and AM Interference Signals Against the AN/PRC-10," The University of Michigan, Electronic Defense Group, Tech. Memorandum No. 33 (2262-135-T), December 1956.

71. Lindsay, W. J., W. J. Parker, and D. H. DeVries, "Study of Susceptibility to Jamming of Soviet Type RBM-1 Receiver," The University of Michigan, Electronic Defense Group, Tech. Memorandum No. 36 (2262-140-T), February 1957.

72. Rothschild, D. R. and D. H. DeVries, "An Investigation of Several Jamming Signals Against Voice Communications Systems," The University of Michigan, Electronic Defense Group, Tech. Memorandum No. 37 (2262-146-T), April 1957.

73. Rothschild, D. R. and D. H. DeVries, "Effectiveness of Jamming Signals Against an AM Voice Communication Receiver," The University of Michigan, Electronic Defense Group, Tech. Memorandum No. 51 (2262-172-T), November 1957.

74. Hellwarth, G. A., D. E. Ross, and D. R. Rothschild, "Discussion of Sampled-Speech Phenomena Occurring During Jamming of Voice-Communications Receivers," The University of Michigan, Electronic Defense Group, Tech. Memorandum No. 63 (2262-195-T),) October 1958.

75. Reports produced under Contracts DA-36-039 sc-72738 (1956-1958), DA-36-039 sc-78240 (1955), DA-36-039 sc-73195 (1957), DA-36-039 sc-64442 (1956), General Electronics Laboratories, Cambridge, Massachusetts.

76. Hellwarth, G. A., D. E. Ross, and D. R. Rothschild, A Discussion of Sampled-Speech Phenomena Occurring During Jamming of Voice-Communication Receivers, *Transactions of the 1958 Symposium.*

77. Hellwarth, G. A., D. E. Ross, and D. R. Rothschild, "Jamming of FM Voice Communications Receiver Including the Effect of Receiver Detuning," The University of Michigan, Electronic Defense Group, Tech. Report No. 95 (2899-7-T), September 1959.

78. Electronic Defense Group, Quarterly Progress Report No. 18, Contract No. DA-36-039 SC-63203, The University of Michigan, Period covering July 15 to October 15, 1958.

79. Lindsay, W. J., C. E. Lindahl, and R. H. Dye, "The Mark III Continuous Monitor Jamming System," The University of Michigan, Electronic Defense Group, Tec Report No. 79 (2262-180-T), June 1958.

80. Birdsall, Carlson, Daws, Ristenbatt, Roberts, and Rothschild, "A Study of Pseudo-Random Systems," The University of Michigan, Electronic Defense Group, Tech. Report No. 104, March 1960.

81. Birdsall, T. G. and M. P. Ristenbatt, "Introduction to Linear Shift-Register Generated Sequences," The University of Michigan, Electronic Defense Group, Tech. Report No. 90 (2262-189-T), October 1958.

15

Radar ECM Repeaters and Transponders

W. E. AYER

15.1 Introduction

Since World War II, a large and important group of countermeasures techniques has been developed which is based on the radiation of deceptive signals simulating radar target echoes. In comparison with less sophisticated techniques such as spot and barrage jamming, these newer approaches are very attractive from the standpoint of average power requirement and hence result in equipments which are light in weight and small in size. Repeaters and transponders make use of more of the characteristic information of the signal to be countered and thus can be more efficient in "getting into" the radar receiver involved. In working against pulsed radars, for example, a repeater utilizes the time of arrival of individual pulses to determine the time of transmission of similar echo-like signals. The duty factor of the output stage of the repeater or transponder will thus be comparable with the duty factor of the radar being deceived: "one-way" jammers (spot, swept, barrage) must have a *continuous* power output roughly equal to the *peak* output of the repeater or transponder.

Before proceeding with the discussion, definitions of the terms "repeater" and "transponder" will be given which are descriptive of such arrangements when used in electronic countermeasures.

Repeater: A receiving-transmitting system which, during at least a portion of its program of operation, automatically amplifies and reradiates the signal received at its input.

Transponder: A receiving-transmitting system which does not possess

the "straight-through" feature of the repeater, but depends instead upon some frequency memory or tuning procedure to effect transmission at the approximate frequency of the input signal.

It should be emphasized that these definitions are somewhat arbitrary and are formulated for maximum convenience with regard to the countermeasures art. A particular concession in this direction is involved in the repeater definition. As will be shown later, systems of importance exist which start their "program" as straight-through repeaters and finish in a transponder mode employing frequency memory. By definition, such arrangements will be called repeaters.

The development of effective countermeasures equipments employing repeater and transponder techniques has become possible with the availability of wideband VHF, UHF and microwave amplifiers and oscillators. Devices of fundamental importance to the new techniques include distributed amplifiers, traveling wave tubes (TWTs), backward-wave oscillators (BWOs) or carcinotrons, and voltage-tunable magnetrons. Distributed amplifiers are now quite commonplace and are treated in the literature of References 1 and 2. The various microwave tubes enumerated above are discussed in Chapters 26 and 27.

As indicated previously, the common denominator of all systems in the repeater-transponder category is their use of deception or confusion to counter the radar equipments which they operate against. Very real advantages over so-called "brute force" techniques are available: (1) The low average power output requirements generally result in equipments of modest size, weight, and power consumption. (2) In many situations, the successful operation of the equipment results in the desired nullification of the radar without betraying the fact that countermeasures are being employed. This latter performance characteristic can be extremely important, for it means in effect that the enemy is accepting incorrect information and is being deceived. In the presence of spot or barrage jamming, on the other hand, the enemy realizes that he is being jammed and finds that the information which he seeks is being denied him. He can then attempt to obtain the information he wants by other techniques (radars in other frequency bands, triangulation on the jamming signal, etc.).

To assume from this that all the advantages lie with repeater techniques, however, would be very dangerous indeed. Brute-force methods offer one extremely important advantage not generally shared by repeaters and transponders; it may be described by a single word: universality. A noise-modulated signal of adequate power will obscure a target to all types of radar equipments. A pulsed reply of particular characteristics, on the other hand, may be expected to constitute an effective countermeasure only against a specific class of radars. Equipments employing repeater techniques thus

must have their performance characteristics tailored to perform given functions (though several operating modes may be combined in a single equipment). Although the limitation of nonuniversality makes it necessary to obtain intelligence information to ascertain the characteristics of the enemy radars and restricts the usefulness of a given equipment, the unique advantages of deception and confusion systems, when properly used, make them an increasingly important part of the total countermeasures art.

In Section 15.2, the parameters of the repeater-transponder problem are discussed, showing the relationships between systems gain, operating frequency, and target cross section and radar and jammer output powers, radar-jammer range, and antenna gains. Section 15.3 is devoted to the arrangement of the r-f circuitry of typical repeaters and transponders. Section 15.4 treats the common deception techniques in use today, with a discussion of the general nature of circuit arrangements which will produce the various deceptive modulations.

15.2 Parameters of the Problem

To establish the possibility of deceiving an active system by means of repeater or transponder techniques, the deception system must possess adequate gain and power output capability to provide a signal at the radar receiver input which is at least commensurate with the true echo signal received from the countermeasure-equipped target. The gain and power requirements may be separated conveniently since they depend upon several different parameters.

The total gain requirement is directly determined by three considerations:

1. The effective radar cross section of the countermeasure-equipped target.
2. The jamming-to-echo signal ratio (J/S ratio) required to produce the desired effect at the radar.
3. The frequency of operation.
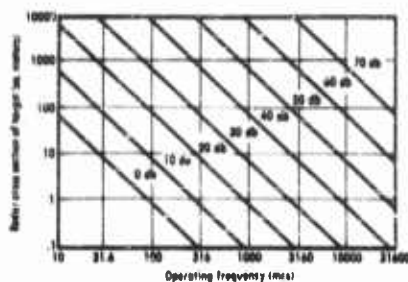
FIGURE 15-1 System Gain Required for Unity J/S Ratio

A plot of the system gain requirement to produce unity J/S ratio at the radar is given in Figure 15-1. The chart is used by locating the intersection of operating frequency and target cross section lines and interpolating the system gain from the adjacent constant-gain lines. Thus for a target cross section of 100 square meters and an operating frequency of 100 megacycles per second, a gain of approximately 62 decibels is required for unity J/S. If a

fairly typical J/S of 10 decibels were desired, the corresponding system gain requirement would be 72 decibels. The total system gain thus determined includes receiving and transmitting antenna gains referred to an isotropic radiator as well as the net "electronic" gain in the repeater proper. Thus, if horn antennas having 10 decibels gain were used for transmitting and receiving at the repeater, the active gain requirement in the above example would be 72 minus 20 or 52 decibels.

The system gain requirement is simply the ratio of the effective radar cross section to be simulated to the capture area of an isotropic radiator. For the unity J/S condition plotted in Figure 15-1, the required power gain is given by

$$G = 1.4 \times 10^{-4} A_s \, (F_{mc})^2 \qquad (15\text{-}1)$$

where A_s is the target cross section in square meters and F is the operating frequency in megacycles per second. The gain requirement as developed in the example above must frequently be considered to be a minimum value which is reached as the radar-to-target range decreases to the smallest separation at which the postulated J/S ratio is to be obtained. In all systems in which saturation (power limiting) is approached or achieved at this closest point of approach, the system gain will decrease with decreasing range. Thus, achievement of the desired gain at minimum range will result in a gain in excess of the required value for large ranges. Allowance must be made for this "excess gain" in all systems depending upon isolating mechanisms to prevent loop oscillation.

The maximum power output that must be developed by the equipment is determined by the following factors:

1) The effective radar cross section of the countermeasures-equipped target.
2) The J/S ratio required to produce the desired effect at the radar.
3) The power gain of the countermeasure transmitting antenna.
4) The effective radiated power (ERP) of the radar to be countered.
5) The minimum range at which the countermeasure is to perform.

Items 1 and 2 determine how large an apparent target the equipment must simulate and are identical to the first two considerations previously stated for gain determination. Item 3 is conveniently thought of as establishing the transmitting antenna cross section with which the desired apparent target size must be simulated. Items 4 and 5, in conjunction with 1, determine the true target echo power with which the countermeasure must compete.

The calculation of the power output requirement may be carried through

as follows: (1) Determine the incident power density at the target by means of the relation

$$\Gamma_{den} = \frac{P_r G_r}{4\pi R^2} \text{ watts per square meter} \qquad (15\text{-}2)$$

where P_r and G_r are the output power in watts, and the antenna power gain of the radar to be countered and R is the radar-to-target range in meters. (2) Find the apparent reradiated (target echo) power by multiplying the power density by the radar cross section (σ) of the target. (This power is that which, if radiated at the target by an isotropic antenna, would produce a signal at the radar equal to that due to reflection from the target.) (3) Find the "electronic" power output (P_t) required of the countermeasure by multiplying the reradiated power from step (2) by the J/S ratio desired and dividing by the gain of the transmitting antenna (G_t) referred to an isotropic radiator. The above series of steps may be combined to give the following expression:

$$P_t = \frac{P_r G_r \sigma}{4\pi R^2 G_t} \cdot \frac{J}{S} \text{ watts} \qquad (15\text{-}3)$$
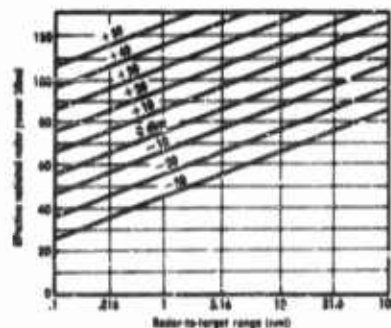
FIGURE 15-2 Effective Radiated Repeater Power for Unity J/S Ratio

Equation (15-3) has been utilized to develop Figure 15-2. In this plot, normalized values of 1 square meter for target cross section and unity J/S ratio have been utilized. The use of the chart is readily understood by following through an example. Assume the following parameters:

$$P_r = \quad 250 \text{ kw } (+54 \text{ dbw})$$
$$G_r = \quad 1250 \text{ (31 db)}$$
$$R = \quad 10 \text{ nautical miles}$$
$$\sigma = \quad 100 \text{ m}^2 \ (+20 \text{ db})$$
$$J/S = \quad 10 \text{ (10 db)}$$
$$G_t = \quad 10 \text{ (10 db)}$$

The combination of P_r and G_r gives an ERP for the radar of $+85$ dbw. An ordinate of this value and an abscissa of 10 miles yields a normalized ERP of -11 dbw. This is converted to equipment power output by adding 20 db for the assumed 100 square meter radar cross section, adding 10 db for the desired J/S ratio, and subtracting 10 db for the countermeasure transmitting antenna gain. The required power output is thus -11 dbw $+$ 20 db $= +9$ dbw or 8 watts.

For the hypothetical situation assumed in the example just carried through, the minimum gain and power output requirements for the equipment have been established at the minimum range at which specified performance is required. With these two quantities determined, the minimum-range input signal may be found directly. The input signal to the repeater is equal to its required power output ($+9$ dbw) less the required electronic gain (52 db). Referring the output power to a milliwatt, $P_{in} = +39$ dbm -52 db $= -13$ dbm for the example cited. As the range is changed, the output power required varies in accordance with $20 \log_{10} (R_2/R_1)$, corresponding to a 20 db change for a 10 to 1 range variation. The corresponding input signal power will vary in the same way with range, whereas the required gain will remain constant.

With all parameters of the problem specified, the minimum gain and minimum output power as determined above may be said to be "consistent." That is, an equipment designed with gain and power output capabilities proportioned as indicated will become gain- and power-limited simultaneously as range decreases to the value utilized in the design. If similar performance is then required at a smaller range or against a more powerful radar, for example, a proportional increase in both the gain and power output of the repeater will be required.

15.3 Basic Radio Frequency Systems

In this section, several of the basic arrangements will be considered by means of block diagrams, followed by a description of the various techniques and devices currently utilized for signal amplification.

15.3.1 Straight-Through Repeaters

The simplest r-f arrangement, conceptually, and one of considerable practical importance, is shown in the block diagram of Figure 15-3. This configuration is referred to as "straight through" because the signal is received, amplified, and reradiated without frequency translation. It is "wide open" in frequency in the sense that it is receptive at all times to input signals anywhere within its operating-frequency range. A fundamental (and frequently difficult) requirement of this arrangement is that the decoupling or isolation between receiving and transmitting antennas must exceed the active gain at all frequencies. If this requirement is not met, the system will oscillate at one or more "natural" frequencies determined by the amplitude and phase characteristics of the complete loop.

FIGURE 15-3 Straight-Through Wide-Open Repeater

15.3.2 Swept Repeaters

For some deception systems, it may be necessary, desirable, or both, to operate on a sweeping narrow-band basis rather than on the continuously wide-open basis discussed above. Thus, if a number of c-w signals are present simultaneously and can be countered effectively by periodic replies from the repeater, this arrangement may be advisable. The peak output capability of the equipment is thereby reduced to that required to handle a single signal, and problems of small-signal suppression are largely eliminated.

In its simplest form, such a system would appear as the straight-through repeater of Figure 15-3, with the addition of a swept narrow-band filter (perhaps a motor-driven cavity) in series with the amplifier chain ahead of the output stage. A more elegant approach, possessing several advan-

FIGURE 15-4 Swept Double Conversion Repeater

tages over the simple swept filter arrangement, is illustrated in Figure 15-4. The incoming signals are first amplified, then translated by means of a mixer and swept local oscillator to a convenient intermediate frequency. Most of the required system gain is achieved by a narrow-band i-f amplifier, after which the signal is translated back to the received frequency by means of a second mixer and the common local oscillator. This amplified signal then drives the broadband

power output stage and the signal is reradiated. In this arrangement, only a small portion of the total gain must be achieved on a broadband basis, simplifying the circuitry. A second advantage, to be discussed in more detail later, is that delay may be conveniently introduced in the i-f section to cause a shift of the transmitted signal frequency with respect to that of the received signal. This effect, which can disrupt the operation of some active systems utilizing doppler shift, results from the change in the output frequency of the swept local oscillator, which occurs during the time required for the signal to pass through the i-f section.

15.3.3 Gated Repeaters

In applications where it is impractical to obtain sufficient isolation between receiving and transmitting antennas to prevent repeater oscillation, or when it is necessary to perform both the receiving and transmitting function with a single antenna, it is necessary to introduce time gating. A simple gated two-antenna repeater is illustrated in Figure 15-5, and its operating program in
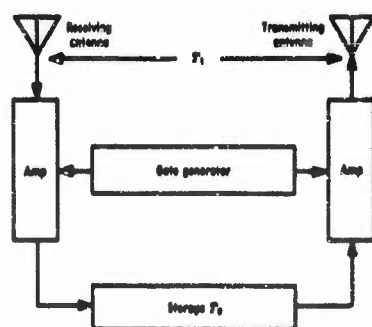
FIGURE 15-5   Simple Gated Repeater
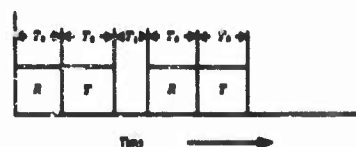
FIGURE 15-6   Program of Gated Repeater

Figure 15-6. Starting at the beginning of a receiving period, the input amplifier is gated "on" and the output amplifier gated "off." The transmitting period starts at the end of the receiving period and lasts for a time T_0, at which time the output amplifier is gated "off." For a period T_1, both output and input amplifiers are held off. This time, frequently described as the "cool-off period," is required to allow direct signals between antennas as well as reflections from nearby objects to fall below the input amplifier threshold. The length of this off period depends upon the characteristics of the particular repeater involved and the environment in which the antenna(s) operate.

Allowance for the loss of on-frequency gain and power resulting from the less-than-unity duty cycle inherent in gated systems is necessary. Usually, the gain and power output must be increased by the factor $(1/\triangle)^2$ over the values determined in Section 15.2, $\triangle$ is the transmitting duty factor, $T_0/(3T_0 + T_1)$. Thus if $\triangle$ is 0.316, "peak" gain and power output must be increased by 10 decibels to realize the desired on-frequency values. This relation holds exactly when the repetition frequency of the gating system is in excess of half the receiving bandwidth of the active system being countered, so that none of the sideband power resulting from the gating process is effective. The $(1/\triangle)^2$ factor is thus simply the ratio of carrier power to peak power for a gated signal. In practice, it is usually found that the relation is correct for equipments designed to work against inherently narrow-band c-w systems, but is slightly pessimistic where active systems employing short pulses are concerned.

15.3.4 Single-Frequency Transponders

The basic arrangement of a simple single-frequency transponder is shown in Figure 15-7. This sort of equipment, in view of its restricted capability, has largely been supplanted by more versatile devices. Operation of the system of Figure 15-7 is as follows: The receiver (typically a superheterodyne) is tuned to the frequency of the radar to be countered to provide a trigger pulse to the transmitter each time a pulse is received. The trans-

FIGURE 15-7 Simple-Single-Frequency Transponder

mitter is similarly tuned to the radar frequency and transmits one or more pulses for each trigger input, the number and timing of the output pulses constituting the deception. A transmit/receive switch (T/ ' amiliar in radar practice is utilized to allow operation of the equipment with a single antenna. This type of device, without deceptive modulation, is in common usage as a radar beacon to identify and extend the detection range of friendly aircraft.

15.3.5 Search-Lock-Jam Transponders*

A much more sophisticated transponder than that shown in Figure 15-7 employs very rapid automatic tuning to provide essentially simultaneous operation against a number of pulsed radars. One form of this arrangement is
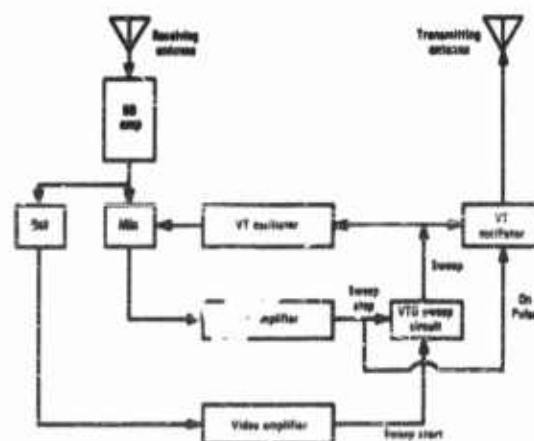
FIGURE 15-8 Elements of Search-Lock-Jam Transponder

illustrated in Figure 15-8. The operation is as follows: The leading edge of a received pulse is detected and amplified by a wideband video amplifier and is used to initiate the sweep on a voltage-tunable oscillator. The output of this oscillator is applied as a local oscillator signal to a mixer which has the received signal applied to its input. The local oscillator sweeps rapidly over the operating frequency range of the equipment until the difference-frequency output of the mixer is within the bandpass of the connected i-f amplifier. The i-f output signal is then used to stop the voltage-tunable local oscillator and the frequency-determining voltage is clamped at this value for a predetermined period of perhaps several tens of microseconds. When sweep-stopping occurs, a second high-powered voltage-tunable oscillator is

*See Reference 3.

gated on, operating as the transmitter during the "stop" period. The transmitted frequency is arranged to approximate closely the received signal frequency through appropriate modification of the applied stopped-sweep voltage to compensate for the i-f offset and any differences in voltage-frequency characteristics of the two oscillators.

The signal acquisition process just described must be completed during a single pulse from the radar. Since the radar pulse may last one microsecond or less, large bandwidths, small delay times, and very fast sweep rates are mandatory. Techniques of this sort were not feasible until the advent of electronically tunable oscillators and are now successful only when careful attention is given to circuit design. To the extent that transmission of the required deceptive signals can be accomplished within a few tens of microseconds following a given received pulse, an equipment of this type can be effective against several (possibly 10 or more) radars which are "looking" at the target simultaneously. This multiple-signal capability results from the low duty factor (typically 0.001) employed by pulsed radars, and the random times at which pulses from a number of unsynchronized radars will arrive.

15.5.6 Low-Frequency Amplification*

The principal method for obtaining an appreciable output power for broad-band operation below approximately 400 megacycles per second employs a number of beam tetrodes in a distributed amplifier configuration. For example, six 4X250B tubes (250 watt dissipation tetrodes) will provide a c-w output power of 100 watts and a power gain of 10 decibels over the 10 to 300 megacycle per second range. Low duty-cycle pulse operation of such an arrangement will yield peak output powers in excess of a kilowatt. A second method of obtaining relatively high output powers in this portion of the spectrum employs a number of narrow-band amplifiers each of which covers a small portion of the total frequency band. Inputs and outputs of the several stages are paralleled by means of frequency-selective networks. This arrangement has more complicated circuits than a distributed amplifier, but has the advantage that several signals at different frequencies may be handled simultaneously with less trouble from small-signal suppression.

Low-level gain is obtained in this frequency range either with distributed or cascaded amplifiers. Distributed amplifiers are simple and tend to be quite reliable since the failure of a tube or two in a multitube stage will result only in a nominal decrease in gain. Amplifiers composed of a number of cascaded stages will generally require complex interstage networks to realize a suitably flat over-all gain characteristic, but will usually use a con-

*See References 1 and 2.

siderably smaller number of tubes to obtain a given gain and bandwidth. The choice between the two approaches can be made only by considering all factors of a specific application.

15.3.7 High-Frequency Amplification

All wideband repeaters above 400 or 500 megacycles currently depend upon TWTs both for low-level gain and output power. It appears likely that development of negative grid tubes specifically suited to distributed amplifiers will extend the upper frequency limit of this technique to 1000 megacycles within the next few years, but coverage of the remainder of the microwave spectrum is likely to depend entirely on TWTs, for the foreseeable future. A large number of different TWT types are now available, with much progress being made in providing increased ruggedness, longer life, lighter weight, higher power outputs, and lower noise figures.

A typical low-level tube covers an octave in frequency anywhere between the limits of 500 and 12,000 megacycles, provides 30 decibels of gain over its band, and has a saturation (maximum) power output of several tens of milliwatts. Tubes with these characteristics are available with permanent magnet focusing, weigh less than 5 pounds, and require only about 10 watts total power input. Similar tubes are available which provide an output of about one watt and require about 50 watts total input power.

In the higher power ranges, bandwidths of 1.5 to 1 are typical. Powers of 100 to 200 watts c-w can be obtained from solenoid-focused tubes exhibiting 20 decibels of gain. Pulsed powers in excess of 1 kw at 30 decibels of gain are common for solenoid focused tubes, and some tubes of these capabilities have been made with permanent magnet focusing.

15.3.8 Electronically Tunable Devices

A number of developments in recent years have led to the capability for rapid electronic tuning. For the lower frequencies, the principal tool is a variable inductor which is tuned by changing the amount of direct current passed through a control winding. Devices of this type may be employed as the tuning elements of amplifier and oscillator circuits. Operation is practical up to several hundred megacycles, and bandwidths of an octave may be realized. High-power models can be employed to obtain about 100 w. of output power. Voltage-tunable capacitors are available and offer promise for use in place of (or in conjunction with) current-tunable inductors. An important inherent advantage of the tunable capacitor is its lower "inertia," resulting in higher tuning rates for a given tuning power.

In the microwave region, the most important electronically tunable device is the backward-wave oscillator (BWO). This member of the TWT family

is available in a wide variety of tuning ranges and power outputs. Bandwidths are typically 1.5 or 2 to 1, and tubes are available covering all principal radar bands. Tubes for operation at frequencies in excess of 50 kmc have also been built. Continuous output powers of several hundred watts are obtainable at frequencies below X-band; low-level tubes suitable for local oscillator service are available to provide continuous coverage to about 40 kmc. Both solenoid and permanent magnet focusing are used at all power levels, with current developments favoring the lighter and more efficient permanent magnet structures.

A promising offshoot of the BWO is the backward-wave amplifier (BWA). These devices have the same electronic tunability and other general characteristics as the BWO, but function as amplifiers to provide upwards of 30 decibels gain at the tuned frequency. Passbands for a given tuning voltage are typically a few tenths of a percent up to one or two percent of the tuned frequency, depending on the adjustment of the beam current.

Probably the most impressive feature of backward-wave tubes from the countermeasures point of view is the extremely rapid tuning rates which may be employed. At S-band, for example, typical tubes will oscillate continuously while being swept at rates as high as 10,000 megacycles per second per second. The search-lock-jam arrangement depicted in Figure 15-8 requires tuning rates of the order of 20 percent of this maximum capability.

15.4 Deception Techniques

In this section, a number of deception techniques will be discussed with emphasis upon how a desired effect may be accomplished with a given r-f system arrangement. Although it is not practical to consider all of the possible combinations of r-f systems and deception techniques in detail, it is believed that the examples used here will allow the reader to assemble for himself other arrangements from the "building blocks" provided.

15.4.1 False Target Generation

The most important deception technique against search and ground-controlled intercept (GCI) radars consists of the the generation of a number of false targets. This technique has two objectives: (1) The enemy will be unable to ascertain which of the targets he sees are real, and hence will have difficulty in taking the appropriate action. (2) He will believe that a massive attack is being made from a given quarter and will commit his retaliatory powers to counter this attack, while the real striking force is approaching slightly later from a different direction.

A multiplicity of false targets which will appear at constant range but over a large azimuth sector can be obtained with a straight-through repeater. To prevent the identification of the real target by comparing echo amplitude, some sort

of automatic gain control (agc) is required in the repeater. One arrange-

FIGURE 15-9 False-Target Repeater

ment is indicated in Figure 15-9. The detector and the agc amplifier are designed to cause the repeater output to vary inversely with the strength of the received signal on a pulse-by-pulse basis so that little or no output is obtained when the main beam of the radar "looks" at the target, whereas maximum output power is radiated when the illumination by the radar is at a low level. Since a prohibitively large dynamic range would be required in the repeater to prevent at least some azimuthal variation in the false target amplitude, which would be symmetrical about the true target position, an additional subterfuge is desirable. One approach is to introduce a random gain modulation or gating in the repeater to destroy the symmetry of the radar display. This is typically a "slow" modulation which changes the output signal amplitude appreciably over a time corresponding to several hundred radar pulse periods. A second superimposed modulation is desirable to increase the realism of the false targets. This is again a random modulation, which changes the output signal on a pulse-to-pulse basis to simulate the normal target scintillation. In a wideband microwave false-target repeater, all three of the "modulations" discussed above might be combined and applied to the control grid of the output TWT.

The false targets produced by the system just described can be satisfactorily "real" in all respects but one: they all appear at the same range from the radar. As such, the radar may be unable to select the true target for retaliation, but it is extremely unlikely that the radar's position is the focal point of a precise radial attack. To add the missing dimension of variable range to the false target picture, some sort of "frequency memory" must be incorporated. The single-frequency transponder of Figure 15-7 possesses this memory in the form of the fixed tuning of the receiver and transmitter and can be used satisfactorily against a single, fixed-frequency radar. The search-lock-jam configuration in Figure 15-8 can accomplish the same result against a number of fixed- or variable-frequency radars. To "show" the radar a number of radially aligned targets, a fixed gating program would be provided giving the desired target spacing. The three modulations discussed in connection with the false-target repeater can be added to the search-lock-jam transponder to create a wide variety of effects. The gating program can be altered on a random basis or as a function of the signal power received from the radar to produce clusters of targets at various ranges and azimuths.

15.4.2 Range Gate Pull-Off[*]

A deception which has been found to be quite effective against pulse tracking radars involves "stealing" the timing gate which during normal operation of the radar is accurately centered at the target echo range. The video signal which supplies agc to the radar receiver and controls the angle-tracking functions is taken through this gate, so that a deception which causes this gate to lose the target will disrupt the tracking function. A repeater with frequency memory or a search-lock-jam transponder arrangement may be used to perform range gate stealing. (Frequency memories employing several hundred feet of delay cable and a TWT to operate as a multimode memory have been used successfully in microwave gate-stealing repeaters. This type of memory is discussed more fully in Chapter 24). The program of a range gate pull-off cycle is as follows: Initially, the countermeasures equipment repeats the pulses received from the radar with minimum delay. If the deception is to be successful, this minimum delay must usually be small in comparison to the radar pulse length, and the peak output power must provide a reradiated power density in excess of that due to the true target echo. Typically, J/S ratios of 10 decibels or more are sought. With the above conditions fulfilled, the countermeasure has "captured" the gate and the reply is slowly delayed with respect to the received signal (and hence the true target echo) until a total delay of 2 to 10 microseconds is achieved. Typical pull-off rates are of the order of 1 microsecond per second. With the range gate pulled off the target, the delay is snapped back to a minimum value. Since the radar range gate is unable to follow the fast recycling, it is left at the range corresponding to the maximum pull-off, where no target echo is present, so that range tracking is lost. If the radar has a range search function, it may reacquire either the target or the false pull-off signal on its next cycle. If the pull-off operation is continually repeated (at a constant rate or preferably with some randomness of rate and period), the radar will operate for only a small portion of the time with true target information, and the tracking accuracy will be seriously degraded. If the radar is manned and has provision for manual override of the automatic range tracking operation, a skilled operator may be able to increase substantially the percentage of time that the radar operates on true target data. His job is much more difficult and fatiguing if random programming of the pull-off is employed.

The circuit arrangement of a typical range gate pull-off repeater is shown in Figure 15-10. In the absence of an input signal, the r-f switch in the TWT/delay cable loop is open and the input tube functions as a straight amplifier. When a signal is received, it is detected and causes the program

[*]See Reference 4.

unit to close the memory loop and simultaneously pulse on the output TWT

FIGURE 15-10 Range Gate Pull-Off Repeater

with a self-generated pulse which is approximately the same length as that employed by the radars to be countered. As successive pulses are received, the memory loop is repeatedly closed, each time for a period equal to the maximum delay which will be used in the pull-off cycle. Meanwhile, the on-pulses applied to the output tube are progressively delayed in time until the maximum offset is reached. At the conclusion of a cycle of operation, the pulses applied to the output tube revert to the minimum delay time and the program is repeated. Randomness of the pull-off operation, if employed, is controlled by the program unit.

15.4.8 Scan-Rate Modulation*

A second deception which may be used against tracking radars employing conical-scan angle tracking utilizes modulation of the repeater or transponder output at or near the conical-scan frequency employed by the radar. A particular form of this technique, which has proved to be very effective, uses square-wave (on-off) gating of the output stage of a repeater at a variable frequency which includes the scan frequency of the radar being countered. As this modulation approaches the radar scan frequency, increasingly large error signals appear in the radar servo loops, causing rapid random gyration of the antenna system. Frequently, the transients thus introduced are sufficient to cause the radar to lose the target completely. Since the scan channel passbands of radars of this type are quite narrow (typically 1 to 4 cycles), the modulation in the countermeasure must be swept slowly if a maximum effect is to be realized. Against typical ground-based equipments, sweep rates of the order of 1 or 2 cycles per second per second give optimum results. The low sweep rates which must be used in this technique point up its principal weakness: a considerable amount of time is required to cover the range of scan rates likely to be employed by the enemy, so that disruptive effects will be introduced in any one equipment at relatively infrequent intervals. The effectiveness of this technique (particularly when combined with range-gate pull-off as discussed later) is so great, however, that even its infrequent application may be adequate to render the radar useless.

*See Reference 4.

15.4.4 Inverse-Gain Modulation*

As an alternative to the noncoherent scan-rate modulation, as discussed above, the modulation envelope of the tracking signal may be detected, inverted, and applied to the countermeasures system to present an inverted error signal to the radar. This approach depends upon the fact that all conical-scan systems operate only by virtue of angular error signals derived from the target echo, and hence there will always be some conical-scan modulation of the pulses received at the target. To be most effective, a deceptive repeater of this type is gated on only during the minimum 180 degree segments of the received modulation envelope. Rather high J/S (20-30 db) are required to introduce angular errors at the radar in excess of a beamwidth. The power output required from the jammer is quite modest, however, since the large gain necessary to realize an adequate J/S ratio is "active" only during minimums of the received pulse train. The virtue of the inverse modulation scheme as compared to swept scan-rate modulation is that it works continuously at the scan rate of the applied signal and can be made to operate effectively against a wide range of conical-scan modulation frequencies. A principal disadvantage of this technique is the difficulty encountered in devising a system that can work effectively against a number of radars simultaneously. Arrangements capable of simultaneous operation with good efficiency against two or three radars have been built, but beyond this, the complexity appears to outweigh the results obtainable.

15.4.5 Combined Range-Angle Deception†

The effectiveness of the angle modulation schemes just discussed may be increased (in some situations by a large factor) by first pulling the range gate away from the target echo and then applying the angle deception. The advantage of this procedure is that the deceptive angle modulation, when applied, is not in competition with the target echo signal which yields true angle error information. This combination technique is particularly desirable for increasing the disruption of the radar function when inverse gain is employed since this type of angle deception is characteristically less violent in its effect on the radar when target competition is present, with much less tendency to break track.

15.4.6 Velocity-Gate Pull-Off‡

Active radar systems of the c-w doppler type may be countered by stealing the velocity gate of the locked-on system. This is analogous to range-

*See References 5, 6, and 7.
†See Reference 4.
‡See Reference 6.

gate stealing in pulsed (time) systems. Here gating is in frequency, so that a frequency shift of the output signal with respect to the received signal is required to accomplish the same sort of effect. The frequency shift operation is accomplished in microwave repeaters through "serrodyne" (sawtooth) modulation of the direct voltage applied to a TW tube helix. This technique,* which is discussed in Chapter 24, provides an output signal which is displaced from the input signal by the frequency of the applied sawtooth. The displacement will be upward with respect to the input signal if the sawtooth voltage is positive (the sloping portion of the waveform increasing with time). Conversely, a negative-going sawtooth will result in a downward displacement of the output frequency.

By changing the frequency of the sawtooth modulation on a programmed basis, a velocity-gate pull-off may be accomplished. Since gate widths in radar systems of this type are quite narrow, it is necessary to initiate the program with an offset no greater than a few hundred cycles. Assuming an adequate J/S ratio (perhaps 10 db), the velocity gate may be stolen by increasing the sawtooth frequency at a rate commensurate with the velocity tracking rate capability of the system being countered. When maximum desired offset has been achieved (values of 5 to 50 kcs are typical), the sawtooth may be "snapped back" to the minimum value and the cycle repeated. When working against certain types of equipments, it may be desirable to hold the offset at a maximum value for a short time while additional modulation such as AM noise is applied. Total pull-off periods of the order of 1 to 10 seconds are typical for operation against most c-w systems in use at present. It is quite often advantageous to use a down-frequency pull-off in preference to an up-frequency one. This preference results from the fact that some c-w homing missiles utilize a drop in doppler to arm or detonate the warhead.

At VHF and UHF frequencies, a circuit arrangement of the sort depicted in Figure 15-4 can be utilized to produce frequency offset of the output signal. A variation of the amount of delay in the i-f section can be used to change the amount of offset. Convenient techniques for rapid delay variation are not available, and the alternative approach of changing the oscillator sweep rate is more practical. A serious difficulty with all techniques involving sweeping is the reduction in on-frequency power due to the low effective duty cycle. Another technique which has been proposed for the generation of frequency-shifted signals at the lower radio frequencies and which appears suitable for broadband operation is as follows: The signal to be shifted is applied to a phase-shift network which has four outputs with, respectively, 0°,

*See Reference 8.

90°, 180°, and 270° shift to its input. Each of these outputs feeds a separate gated amplifier, the outputs of the gated amplifiers being connected in parallel. A four-output gate generator would be provided to gate on the amplifiers sequentially to produce the effect of a rotating phase of the output signal. The sampling rate would determine the amount of offset, whereas the "direction" of the sampling would establish the up or down direction of the shift.

REFERENCES

1. Ginzton, E. L., W. R. Hewlett, J. H. Jasberg, and J. D. Noe, Distributed Amplification, *Proc. IRE*, Vol. 36, pp. 956-969, August, 1945.

2. Horton, W. R., J. H. Jasberg, and J. D. Noe, Distributed Amplification: Practical Considerations and Experimental Results, *Proc. IRE*, Vol. 38, pp. 748-753, July, 1950.

3. Kincheloe, W. R. Jr., and F. M. Turner, "The S-601 Micro-Sweep Receiver," Stanford Systems Techniques Laboratory Report #601-1, 25 June 1958. (CONFIDENTIAL.)

4. Ayer, W. E., D. J. Grace, J. L. Grigsby, and M. Wright, "The S440 Repeater Jammer," Stanford Systems Techniques Laboratory Report #440-1. 30 April 1957. (SECRET)

5. Brandenburg, R. L. (Classified Title), NRL Report #4811, 27 September 1956. (SECRET)

6. Wright, M., and W. Ayer, "A C-W Deception Jammer," Stanford Electonics Laboratories Report #441-1, 31 December 1957. (SECRET)

7. Cosby, L. A., "The ALQ-H(X) Countermeasures System," NRL Report #5157, 16 July 1958. (SECRET)

8. Cumming, R. C., "Frequency Translation by Modulation of Transit Time Devices," Stanford Electronics Laboratories Report No. 39, August 1, 1955.

16

Fuze and Communications Repeaters

B. F. BARTON

16.1 Introduction

Considerable effort has been made in the United States to develop repeater jammers for use against radio doppler proximity fuzes. Devices with some of the attributes of a repeater have also been constructed for use against communication links. In spite of their common frequency range (below 500 megacycles), typical devices for the two applications have substantial dissimilarity. The dissimilarity results from different design objectives in the two instances. Fuze repeaters have been built which simply provide, for certain fuzes, a successful imitation at the fuze of the normal return signal. On the other hand, the effectiveness of the communication jamming devices often depends on confusion or masking.

In comparison with other fuze countermeasures equipments, repeaters have a high potential utility against broad classes of radio doppler types. The merit of the simple repeater* as a communications jammer is not clear. In fact, one finds that the merit of repeater-like communication jammers usually depends strongly on their variations from the simple repeater. The emphasis on *fuze* repeaters in the following discussion reflects this appraisal.

16.2 Radio Doppler Proximity Fuze Principles

A discussion of the use of repeaters against radio proximity doppler fuzes

---

*By simple repeater is meant a device which can provide a continuous amplified reflection of a received signal, or a device such as a short-delay repeater which appears to do this in operation against certain targets.

is conveniently introduced with a brief discussion of the single-channel radio doppler fuze. A multitude of signals may be radiated by such a fuze; these signals produce a voltage at some reference point of the fuze which is often easily described in the form

$$c_i = A(t) \cos [\omega_0 t + \phi(t)],\qquad(16\text{-}1)$$

where ω_0 is the radian carrier frequency. Examples of single-channel fuzes which have been studied in the U.S. include:

1. *CW fuze* $[A(t)$ constant; $\phi(t)$ constant$]$
2. *FM fuze* $[A(t)$ constant; for sinusoidal modulation waveform $\phi(t) = \triangle\omega \sin(\omega_m t + \phi_0)]$
3. *PD (pulse doppler) fuze* $[A(t)$ a pulse waveform; in the idealized case $\phi(t)$ is constant within each pulse, and there is essentially no correlation of r-f phase of successive pulses$]$
4. *Noise fuze* $[A(t)$ and $\phi(t)$ are stochastic variables$]$

It is useful to regard the doppler fuze as a practical correlator. Recall the definition of the cross-correlation $\psi_{12}(\triangle t)$ of the functions $e_i(t)$ and $f(t)$:

$$\psi_{12}(\triangle t) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} e_i(t) f(t - \triangle t)\, dt$$

Similarly, the autocorrelation $\psi_{11}(\triangle t)$ of $e_i(t)$ is defined as

$$\psi_{11}(\triangle t) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} e_i(t) e_i(t - \triangle t)\, dt$$

The latter is suggestive of the normal operation of a fuze radiating a signal $e_i(t)$, in that the product of $e_i(t)$ and a delayed return signal proportional to $e_i(t - \triangle t)$ is obtained and then fed through a narrow-band filter in doppler fuzes (this process is approximately equivalent to integrating and averaging). Fuze operation differs from true autocorrelation, however, in that $\triangle t$ is actually varied slowly, so that only a limited time is spent in any increment of delay. This is true of practical correlators in general since an infinite integration time is impractical. Cross-correlation is, in a similar way, suggestive of the fuze processing when an externally generated signal produces a voltage $f(t)$ at the reference point.

It is informative to consider the generation of a triggering signal in a

doppler fuze in some detail. The return signal $e_r(t)$ for a fuze radiating a signal given by Eq (16-1) is assumed to be well represented by the expression*

$$e_r(t) = k_0 e_t(t - \triangle t) = k_0 A(t - \triangle t) \cos [\omega_0(t - \triangle t) + \phi(t - \triangle t)]. \quad (16\text{-}2)$$

Making the further assumption of uniform relative motion between the fuze and the reflector (i.e., $\triangle t = \triangle t_0 - (\omega_d/\omega_0)t$, where ω_d is the doppler frequency and $\triangle t_0$ is a const??), Eq (16-2) becomes

$$e_r(t) = k_1 A \left[t \left(1 - \frac{\omega_d}{\omega_0} \right) - \triangle t_0 \right] \cos \left\{ \left(\omega_0 + \omega_d \right) t - \omega_0 \triangle t_0 \right. $$
$$\left. + \phi \left[t \left(1 - \frac{\omega_d}{\omega_0} \right) - \triangle t_0 \right] \right\}$$

Here, the parameter k_0 has been incorporated in the new parameter k_1, which also characterizes the multiplier gain.

In normal operation, a triggering signal is filtered from the output of a multiplier circuit, which forms the product $e_r(t) \, e_t(t)$. The product $e_r(t) \, e_t(t)$ can be written in the form

$$\frac{1}{2} kA(t)A \left[t \left(1 - \frac{\omega_d}{\omega_0} \right) - \triangle t_0 \right] \cos \left\{ 2\omega_0 t + \omega_d t - \omega_0 \triangle t_0 + \phi(t) \right. $$
$$\left. + \phi \left[t \left(1 - \frac{\omega_d}{\omega_0} \right) - \triangle t_0 \right] \right\} + \frac{1}{2} kA(t)A \left[t \left(1 - \frac{\omega_d}{\omega_0} \right) - \triangle t_0 \right]$$
$$\cdot \cos \left\{ \omega_d t - \omega_0 \triangle t_0 + \phi \left[t \left(1 - \frac{\omega_d}{\omega_0} \right) - \triangle t_0 \right] - \phi t \right\}$$

where k incorporates k_1 and a filter gain cons???. The former of these additive terms is at approximately double the carrier frequency for a narrowband assumption; that is, for

$$\omega_d \ll \omega_0$$
$$A(t) < M \qquad \text{for all } t$$
$$\left| \frac{dA(t)}{dt} \right| < \omega_0 M \qquad \text{for all } t$$
$$\left| \frac{d\omega}{dt} \right| \ll \omega_0 \qquad \text{for all } t$$

where M is a constant and $\epsilon \ll 1$. The signal represented by this term is

*The time variation of the amplitude of the return signal due to time variation of path factor is to be described by the factor k_0.

normally filtered (averaged) out; the triggering signal is in fact selected, with filters and amplifiers, from the multiplier output

$$\overline{e_r(t)\, e_i(t)} = \frac{1}{2} k A(t) A \left[t \left(1 - \frac{\omega_d}{\omega_0} \right) - \triangle t_0 \right]$$

$$\cdot \cos \left\{ \omega_d t - \omega_0 \triangle t_0 + \phi \left[t \left(1 - \frac{\omega_d}{\omega_0} \right) - \triangle t_0 \right] - \phi(t) \right\} \quad (16\text{-}3)$$

Bandpass amplifiers passing a frequency band around the doppler frequency ω_d are employed in CW, PD, and noise fuzes. Near ω_d the energy density in the spectrum of $\overline{e_r(t)\, e_i(t)}$ is high for these fuzes.

The spectrum of $\overline{e_r(t)\, e_i(t)}$ may have several maxima of energy density for other possible radiated signals, such as that of the FM fuze. In the specific case of the FM-by-sine wave fuze, $\overline{e_r(t)\, e_i(t)}$ contains (ideally) discrete components at $n\omega_m \pm \omega_d$ ($n = 0, 1, 2, \dots$), where ω_m is the (radian) modulation frequency. An FM fuze in which the amplifier passes the components at $n\omega_m + \omega_d$ and $n\omega_m - \omega_d$ is termed a J_n fuze. A number of possible triggering signals, each of which has its peculiar time variation of amplitude, are available by choice of n.

The term of $\overline{e_r(t)\, e_i(t)}$ giving rise to the spectral concentration near ω is easily obtained for idealizations of the radiated signals listed at the beginning of this section. Specifically, Eq (16-3) may be written in the form

$$e_r(t)\, e_i(t) = \frac{1}{2} k A(t) A \left[t \left(1 - \frac{\omega_d}{\omega_0} \right) - \triangle t_0 \right]$$

$$\cdot \cos \left(\omega_d t - \omega_0 \triangle t_0 + a(t) \cos (\omega_k t + \theta_k) \right)$$

where

$$\phi \left[t \left(1 - \frac{\omega_d}{\omega_0} \right) - \triangle t_0 \right] - \phi(t) = a(t) \cos (\omega_k t + \theta_k)$$

Now, for the usual fuze radiations, it is found that the time variation (if any) of $a(t)$ is very slow, so that $a(t)$ may be regarded as essentially constant for a few periods at the frequency ω. If the time variation $A(t)$ is also sufficiently slow, it is observed that $\overline{e_r(t)\, e_i(t)}$ is essentially an FM signal with carrier frequency ω_d and sidebands separated by some frequency ω_k. These components are then regarded as varying slowly in amplitude with $kA(t)A(t - \triangle t_0)$, and perhaps with a slow variation of the modulation index $a(t)$. To a good approximation in the usual case, then, the resultant spectrum contains

lines at $\omega_d \pm n\omega_k$, which vary very slowly in amplitude as $kA(t)A(t - \Delta t_0)$ $J_n[a(t)]$. The true doppler term of this function is the carrier component, the amplitude of which is taken as $A_d(\Delta t)$ and which varies with $kA(t)$ $\Delta(t-\Delta t_0) J_0[a(t)]$. The variation with time of the amplitude of the several components is thus dependent on n. In the specific case of the FM-by-sinusoid fuze radiation, it is readily shown that Eq (16-3) has the form

$$e_r(t)\, e_i(t) = \frac{1}{2}\, kc^2 \cos\left\{ \omega t - \omega_0\Delta t_0 + \frac{2\Delta\omega}{\omega_m} \right.$$

$$\left. \cdot \sin\left(\omega t\, \frac{\omega_m}{2\omega_0} - \frac{\omega_m\,\Delta t_0}{2} \right) \cos\left[\omega_m t\left(1 - \frac{\omega}{2\omega_0} \right) + \frac{\omega_m\Delta t_0}{2} \right] \right\}$$

In this case, $A(t)$ is a constant, c, and

$$a(t) = \frac{2\Delta\omega}{\omega_m} \sin\left(\frac{\omega_m}{2\omega_0} - \frac{\omega_m\,\Delta t_0}{2} \right)$$

$$\omega_k = \omega_m\left(1 - \frac{\omega}{2\omega_0} \right)$$

$$\theta_k = \frac{\omega_m\,\Delta t_0}{2}$$

The amplitudes of the doppler components $A_d(\Delta t)$ are given in Table 16-I for several of the fuze types previously listed. These $A_d(\Delta t)$ depict the idealized

TABLE 16-I.   Doppler Component Amplitudes

Fuze Type	Magnitude of Doppler Component $A_d(\Delta t)$	Special Assumptions
CW	k	
J_0-FM	$\frac{1}{2}\, kc\, J_0\left| \frac{2\Delta\omega}{\omega_m} \sin \frac{\omega_m \Delta t}{2} \right|$	$A(t)$ is a constant c
PD	$\frac{k}{T}\, c\, e^{-a^2 t^2} \int_0^{\cdot\cdot} e^{-\cdots} d\nu$	Pulse envelope $A(t) = A_0 e^{-a^2 t^2}$ for $mT < t < mT + \tau$, m an integer T squeezing period
Noise	Envelope of $k\, \frac{\sin \pi \cdots t}{\pi \cdots t}$ (minor lobes fall off inversely with t)	Rectangular noise spectrum of width $\Delta\omega$

time variation of the amplitude of the doppler signal, *apart from variations due to path factor*, antenna pattern, and scintillations of the reflector. The assumptions made in each case are also listed.* The $A_d(\triangle t)$ are sketched
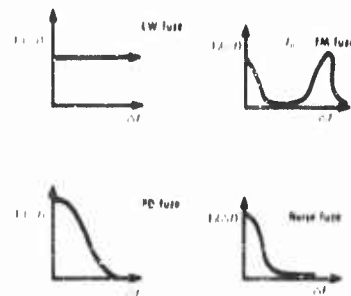
FIGURE 16-1 Sketch of A_d ($\triangle t$) as a function of $\triangle t$

in Figure 16-1 for likely fuze parameters. The A_d are plotted versus delay $\triangle t$ (which is linear with t for our assumptions), a maximum of A_d occurring at $\triangle t$ = 0 for the examples given.† When $A_d(\triangle t)$ tends to remain small for $\triangle t$ larger than some value, the associated $e_i(t)$ is said to give a range-cutoff characteristic (e.g., the PD and noise fuzes above).

The preceding discussion dealt with the basic operation of single-channel doppler fuzes. Actual fuzes are very likely to have a number of properties which result from special techniques such as those listed below. Such techniques can only be mentioned briefly here, and the mention should be regarded as suggestive of the variety of possibilities rather than as comprehensive (Reference 1).

1) CVT (controlled variable-time) action
2) Point-contact detonation capability
3) Special amplifiers
4) Multiple-channels

The CVT firing is widely used in modern U.S. fuzes. The fuze activation is delayed in such a fuze by a mechanical timer for an interval after projection which can be preset prior to use. The objective is to arm the fuze shortly before the anticipated burst time; the interval of fuze activation is held to about two seconds in the case of artillery rounds. Such a mechanism not only provides protection against premature bursts at the launching site but also limits the time available for electronic countermeasures. Radio proximity fuzes for use against ground targets may also be expected to detonate on contact, such a characteristic providing CCM protection against dudding (as by saturation of the fuze amplifier).

A second special technique involves the use of amplifiers having responses which are tailored to characteristics of the (doppler) triggering signal which can be anticipated in specific applications. For example, one may establish

---

*The A_d contain an arbitrary gain constant k.

†$A_d(\triangle t)$ is proportional to the autocorrelation of the radiated signal.

the approximate rate at which the amplitude of the doppler signal increases with time in an application such as a bomb drop. A nonlinear circuit can then be employed in the amplifier, which discriminates against signals having growth rates outside the expected range of values. The use of an amplifier requiring such special properties of the triggering signal is intended primarily as a CCM feature.

Two (or more) of the single-channel fuze circuits such as those discussed above may be incorporated into a multiple-channel fuze. It is possible, by proper interconnection of such channels, to reduce either dudding due to component failure, or vulnerability to countermeasures, or both.

16.3 Basic Considerations in Using Repeater Jammers Against Fuzes

The original objective in using repeaters against proximity fuzes was to simulate a passive reflector proximate to the fuze with an active device located perhaps remote from the fuze. In implementing this scheme, however, a basic problem arises from the fact that substantial gain may be required within the device to offset the large path factor loss resulting from the remoteness of the repeater from the fuze. A jammer with useful gain which employed a single antenna for both reception and transmission would oscillate if operated continuously. The super-regenerative repeater, which will be discussed subsequently, is just such a device.

As alternative solutions, one may have recourse to either a two-antenna system or a system which employs a memory device and time-shares between a receiving and a transmitting mode. Disadvantages of the two-antenna system include the bulk of two broadband antennas in the fuze frequency range and the antenna isolation problem, which proves particularly difficult over the wide frequency range for which operation is sought. The solution involving time-sharing has been widely employed in U.S. repeaters.

It is shown in Reference 2 that the range r_d at which a fuze will be pre-detonated by either a (matched) single antenna repeater or a repeater with two identical closely spaced (matched) antennas can be expressed in the form

$$r_d = .4 \left| H_{\nu} \, \lambda_r \, G_r \sqrt{\frac{A}{D}} \right|^{\nu} \left| f(\rho) \, f_r(\theta, \phi) \, f_r(\theta, \phi) \right| \qquad (16\text{-}4)$$

Here

H_M = "Michigan Height" = the maximum height above an infinitely conducting ground at which the fuze will function when operating without interference.

λ_r = mean wavelength of radiation

G_j = gain of repeater antenna relative to isotropic radiator

A = repeater power gain

D = discrimination factor $= (E_j/E_f)^2$ = the square of the ratio of the jamming field strength at the fuze necessary to detonate the fuze to the field strength of the normal return signal at the fuze necessary to cause detonation

$f(p)$ = polarization factor between jammer and fuze antennas

$f_j(\theta,\phi)$ = relative pattern factor for jammer antenna in the direction of the fuze $[0 \leqq f_j \leqq 1$ for any (θ,ϕ) pair$]$

$f_f(\theta,\phi)$ = relative pattern factor for fuze antenna in the direction of the repeater. $[0 \leqq f_f \leqq 1$ for any (θ,ϕ) pair$]$

In Eq (16-4) it has been assumed that r_d is proportional to the effective repeater antenna length L_j, for which the substitution $L_j = \lambda_j (R_j G_j/120\pi^2)^{1/2}$ has been made. R_j is the real part of the series antenna impedance. Equation (16-4) indicates the function relationship of jammer range capability to the several parameters of the tactical situation. The parameters H_M, λ_f, and f_f are fuze properties and are therefore independent parameters in a specific jamming attempt, whereas G_j, f_j, and A are properties of the jammer. D-factor, $f(p)$, and the maximum jamming range r_d are mutually controlled factors of the

FIGURE 16-2   Gain-Range Curves, Two-Way Problem

fuze and jammer relationship. Equation (16-4) is plotted in Figure 16-2 for $f f_j f_f = 1$.

Now, let us consider a sample calculation in which a very favorable jamming situation is assumed; i.e., $f(p) f_j(\theta,\phi) f_f(\theta,\phi) = 1$, or aligned jammer and fuze antennas. For

G_j = 1

D = 10 decibels

A = 125 decibels

H_M = 33.3 feet

λ_f = 6 feet

one finds

$$r_d \approx 4600 \text{ feet}$$

One way of increasing the maximum range of the jammer in the above example is by using a directional jammer antenna. How-

FIGURE 16-3 Illustration of a Jammer Protection of a Ground Target Area

ever, the tactical value of the device may be reduced significantly by such a modification. This is illustrated in Figure 16-3, where a jammer is portrayed protecting a ground target from artillery fire. The area protected by the jammer is related in a rather complicated way to a volume in space which is the intersection of three volumes:

1. The volume containing possible shell trajectories, which is a simple tube for localized gun emplacement and target.

2. The volume between the locus of CVT activation of the fuze and the minimum safe height for fuze predetonation, which is essentially a volume bounded by two planes for a simple tube of trajectories and a localized target.

3. The volume in which useful coupling is achieved between the jammer and fuze antennas.

The objective of the above discussion was to suggest that the area of protection can become quite small when substantial jammer antenna gain is utilized.* Furthermore, since many factors of the tactical geometry cannot readily be controlled at the jammer, the locus of this area is not easily characterized in practice. Thus, a careful compromise between repeater antenna gain and range capability must be made to insure tactical utility.

16.4 The Time-Shared Repeater

The basic time-shared single-antenna repeater is illustrated in Figure 16-4. The transmission and reception intervals of a practical ground-based single-

*Allowance must also be made for the time interval after CVT activation before a multichannel jammer undertakes jamming of a specific fuze and for the time needed to process a given fuze.
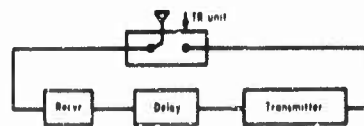
FIGURE 16-4   Time-Shared Single-Antenna Re-
peater

antenna system are normally sep-
arated by an interval of inactivity
a few microseconds long. During
this inactive interval, echoes from
nearby objects which would
otherwise saturate the repeater
input circuit during reception are
permitted to die out. The duty factor of the equipment can be made to
approach one-half by making the length of the repeater delay large compared
to the length of this inactive interval. Sonic (quartz) lines provide large,
compact delays for this application; units with delays of as much as two
milliseconds are available. Transducers have limited these quartz delays to
bandwidths less than about 10 megacycles, and center frequencies to less
than perhaps 80 megacycles per second. The wide frequency band of pos-
sible fuze activity can be covered with such a delay line by using a super-
heterodyne technique. In such a circuit, some portion of the total r-f fre-
quency band is converted to an i-f band in an input mixer. The input mixer
is followed by an i-f section, which contains most of the repeater gain as
well as the delay unit. The output of the i-f section is then reconverted to
the r-f band using a second (output) mixer. The frequency band covered
instantaneously by the repeater can then be varied over the total r-f band
of the equipment by appropriate local oscillator tuning.

A second basic problem of the time-shared single-antenna repeater is the
transmit-receiver switch (TR) unit. One solution to this problem depends
on the use of a distributed amplifier, (DA) as the repeater output amplifier.
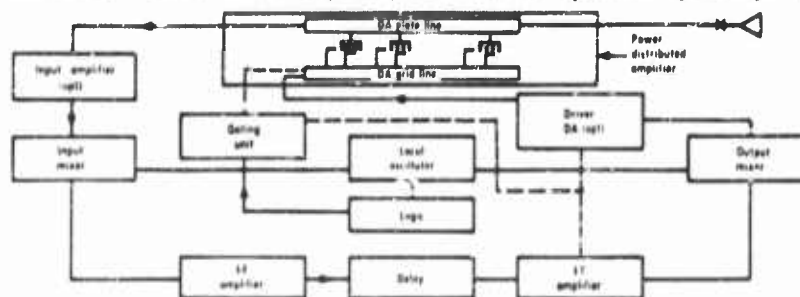
FIGURE 16-5   Block Diagram of a Simple Time-Shared Superheterodyne Repeater

Figure 16-5 is a block diagram of a superheterodyne repeater in which the
output DA is used as the TR unit. During reception, transmission cannot
occur due to the presence of a large negative bias at the control grids of the

power distributed amplifier.* Received signals travel in a reverse direction down the power DA plate line. Transmission is permitted only when an appropriate gating signal is applied at the power DA grid line. During transmission, either the optional input amplifier or the input mixer serves as a reverse termination for the power DA plate line. The reverse termination power of a DA approaches the forward termination (or output) power at certain frequencies, so that substantial power absorption capability is necessary in the input circuit of a high-powered device.

Two basically different patterns of local oscillator frequency variation have been employed in fuze repeaters developed in the United States. The local oscillator has been swept continuously in the case of the short-delay repeater (Reference 3) and stepped in the case of the long-delay repeater (Reference 4). It will be recognized that when the local oscillator of a repeater is swept continuously, as in the short-delay repeater, there is a difference in the mean frequency of the received signal and the mean frequency of the subsequent transmission. Thus, there is what may be referred to as an "artificial doppler shift" of the transmitted signal. This phenomenon may actually be used to advantage since in many tactical situations there is negligible (natural) doppler shift of the signal returned by the repeater. For a device utilizing a 2-microsecond delay line, one finds that an artificial doppler shift of 600 cps results for a local oscillator sweep rate of 300 mc per second per second. With an instantaneous repeater bandwidth of 10 megacycles, repeating of a signal at some given frequency then occurs for about 33 milliseconds. A requirement by the fuze of return signal persistence in excess of 33 milliseconds cannot be met, of course, without lowering the sweep rate of the equipment.

It will be noted that, as the repeater delay is increased for the continuously swept repeater discussed above, the sweep rate must be reduced proportionately in order to keep the artificial doppler shift constant. Thus, as repeater delay is increased, it becomes increasingly difficult to cover the frequency band in which fuze activity is likely within the CVT life of a fuze. On the other hand, the D-factor of the repeater against the CW fuze is nearly proportional to the square of the reciprocal of the transmission duty factor, for delays substantially less than a period at the doppler frequency and for a limited peak repeater output power. In actual practice, the duty factor of a 2-microsecond repeater for use against artillery shells is small (perhaps one-sixth) since the time needed for echoes to die out is often as much as 8 microseconds.

Each component of the fuze radiation produces at the time-shared repeater

*Additional gating may be employed in the transmitter section of the repeater, as shown by the dashed lines of Figure 16-5.

output a signal having a discrete spectrum, the lines of which are separated by the fundamental frequency of the repeater gating waveform. The consequent signal impinging on the fuze is also offset by whatever natural doppler shift is present. In addition, in the case of the short-delay repeater, the impinging signal is offset by the artificial doppler shift due to local oscillator tuning. The fundamental frequency of the gating waveform is relatively high in the case of the short-delay repeater. As a result, in the case of the CW fuze, only the doppler-shifted carrier (natural plus artificial) of the spectrum radiated by the repeater is effective in generating in the fuze a beat signal in the doppler frequency range.

The long-delay repeater covers the total r-f band by stepping the local oscillator frequency and utilizes a delay time T of the order of one-half of the period of the highest doppler frequency generated in normal operation among the fuzes to be countered.* During a jamming attempt against the CW fuze, the local oscillator frequency is maintained constant, producing a return signal carrier differing in frequency by the natural doppler shift resulting from the relative motion of the fuze and jammer. The first sidebands of the signal returned to the fuze by the long-delay repeater again differ from the fuze frequency by either the sum or the difference of the fundamental frequency of the repeater gating waveform and the natural doppler frequency. The novelty of the long-delay repeater lies in the choice of a T (and hence in a choice of a fundamental frequency of the gating waveform) so low that the frequency separation of the sidebands of the return signal is in the doppler frequency region. Which return signal components produce beat frequency components lying within the fuze amplifier passband depends on the passband of the amplifier and the natural doppler shift occurring in a given jamming attempt.†

One very reasonable long-delay repeater design depends on a choice of T such as that mentioned in the previous paragraph. For such a choice it is likely that either the carrier or one of the first sidebands of the return signal will produce a beat signal frequency within the fuze amplifier passband of those target fuzes having the highest doppler frequency in normal operation. Lower beat frequencies (for fuzes having lower doppler frequencies in normal

---

*Long-delay repeater tuning has been accomplished by sequencing among a set of fixed frequency local oscillators. The choice of repeater delay time T is considered further in the next paragraph.

†The beat frequency signal is generated, in a fuze employing an oscillating detector, due to the lock-in phenomenon of oscillators. "Lock-in" refers to the synchronization of an oscillator by an external signal when the external signal is of sufficient strength and is sufficiently close to the undisturbed oscillator frequency. In the "oscillating detector", a detection signal (that is, a beat frequency signal when the normal return or a jamming signal impinges on the fuze) is generated by the nonlinearity of operation of the oscillator tube and can be obtained from some convenient port of the oscillator circuit.

operation) may be obtained simply by not transmitting in certain of the possible transmission periods. These techniques permit effective jamming for the wide range of natural doppler frequencies that might be encountered in tactical situations.

In closing the discussion of superheterodyne fuze repeaters in particular, it is worthwhile to comment that there is a practical advantage in selecting mutually exclusive frequency bands for the r-f, i-f, and local oscillator sections of an equipment. Problems associated with stray leakage are minimized in this way; for example, one reduces the danger that leakage from the local oscillator section may saturate either i-f or r-f amplifiers. However, the loci of both the r-f and i-f bands are constrained to some degree. The r-f band is determined by the equipment application, whereas the i-f band is restricted by the limitations of delay lines. This problem may be solved best by using

a double-conversion technique. Figure 16-6 shows a possible choice of frequency bands for the various sections of a repeater covering the 60 to 300-megacycle band and illustrates the frequency separation principle.

16.5 The Superregenerative Repeater

The superregenerative repeater (Reference 5) is a superregenerative receiver in which the regenerative loop (i.e., the oscillator) is coupled bilaterally to the antenna. The heart of the superregenerative receiver is an oscillator in which oscillations are allowed to build up repetitively. Intermediate to these growth periods, the oscillator is loaded heavily (quenched) in order to dissipate rapidly the r-f energy stored in the oscillator circuit. In the absence of an impinging signal, the r-f pulses (oscillations) build up from circuit noise. When a signal significantly above the noise is present during pulse initiation, however, the oscillations grow essentially from the signal. Thus a small signal which is present at the time of pulse initiation influences both the phase and frequency of the pulse.

FIGURE 16-6 Block Diagram of a Repeater Employing Double Conversion

In the case of the superregenerative receiver the pulses are quenched before the oscillator loop can saturate, for impinging signals in some range

of amplitudes.* The peaks of a succession of pulses tend to follow the AM audio waveform of the incoming signal since the final amplitude of the pulses is closely related to the magnitude of the impinging signal at pulse initiation. Thus, the envelope of the receiver pulses is an approximate reproduction of slow AM on an impinging signal within the dynamic range of the circuit.

The pulses generated in the regenerative loop of the superregenerative repeater are radiated as a result of the bilateral coupling between the loop and the antenna. There is a maximum usable growth period (pulse length) for the superregenerative repeater which results from the existence of radar echoes, and varies from site to site. This limitation arises from the fact that the repeater saturates rapidly when the return energy becomes excessive, the saturation time depends on the terrain characteristics of a given site. The pulse length which may be employed also depends on the prf being employed and the average power limitation of the equipment.

The amplification A of a superregenerative repeater (and receiver) in which complete quenching is achieved between pulses, for very narrow-band signals at levels smaller than that which produces saturation, is a function of the frequency of the applied signal.† This gain dependence is portrayed in Figure 16-7a, and may be contrasted, as shown in Figure 16-7b, with the gain dependence when quenching is incomplete. When the bandwidth of the received signal is significant compared to the width of these gain curves, there may be considerable

(a) Complete quenching   (b) Incomplete quenching

FIGURE 16-7   Gain Versus Frequency Characteristics of Superregenerative Repeaters

difference between the spectra of the received and transmitted signals. In this situation, the spectrum of the signal transmitted by the repeater depends in a relatively complicated way on the exact character of the received signal during the periods of pulse initiation.

Two problems discussed in connection with the time-shared repeater are again encountered with the superregenerative repeater. Successful superregenerative repeater operation against the CW fuze depends on the existence of a doppler shift of the transmitted signal, unless the envelope of its transmitted signal generates an audio waveform at the fuze. It is impractical to maintain an artificial doppler shift (audio) with such a device, so that jam-

*Clearly, the saturation level is related to the dynamic range of such a receiver.

†Amplification refers in this case to the ratio of the magnitude of the input signal to the magnitude of the peak pulse amplitude.

ming effectiveness may depend on relative motion of the jammer and fuze. In addition, the instantaneous bandwidths of known superregenerative repeaters are very small compared to the band of possible fuze activity. Useful sweep rates are restricted by the signal persistence requirements which are likely in modern fuzes. As a result, the r-f band which can be covered effectively with such a jammer is narrow.

16.6 Design Philosophies for General Purpose Repeaters for Use Against Radio Doppler Fuzes

The preceding discussion of repeaters has been confined to their use against CW fuzes. Repeaters are well-suited to this use, and both the time-shared single-antenna repeater and the superregenerative repeater have operated successfully against specific CW fuzes. Conclusions as to the utility of repeaters against doppler fuzes in general are much more difficult. In particular, there are a variety of relatively simple fuzes, the idealizations of which have range-cutoff characteristics. The D-factor is large for repeaters located remotely from fuzes having good range-cutoff characteristics. On the other hand, one expects such range-cutoff characteristics to be continuous. Thus there is a possibility that, for some ranges greater than the design burst height, a *short-delay repeater* signal may have a reasonable D-factor against many doppler fuzes. This suggests that a small short-delay equipment with an omnidirectional antenna might provide reasonably efficient local protection against proximity-fuzed shells. The merit of such an equipment depends in part on whether there is sufficient advantage, in a specific application, to predetonation at a range of a few times the normal burst height.

Alternatively, a jammer design may be predicated on specific expected properties of the enemy fuzes. For example, large repeaters (which m have substantial jamming range to justify themselves) are constructed on the assumption that fuzes with restricted range-cutoff capability will be encountered. Such design assumptions may be justified in certain instances. In particular, a specific property may be characteristic of fuzes for a specific application as a result of volume limitation, economic factors, or the state-of-the-art. However, it is very difficult to successfully predicate the design of equipments with broad utility on fuze properties resulting from specific circuit techniques.

16.7 The Use of Repeater Jammers Against Conventional Voice Communication Links

The merit of the simple repeater as a jammer against conventional voice links has not been proven. The use of a simple repeater to effect a reduction of the field strength at a target receiver is generally impractical, even though

this can be done in principle. Neither can use be made of the fact that a repeater with proper gain and a delay of sufficient length would act as a comb filter to sideband components of the communication signal. The difficulty here is that such filtering may well not greatly reduce message intelligibility. There is even a danger that in using a repeater for communication jamming, the repeater will actually augment the direct transmission. Finally, the periods of transmitter inactivity for the time-shared repeater result in periods during which the communication link is not jammed.*

A multitude of voice communication jammers can be envisioned having some of the properties of a repeater, but having potential merit as a result of basic variations from the simple time-shared repeater. Consider, for
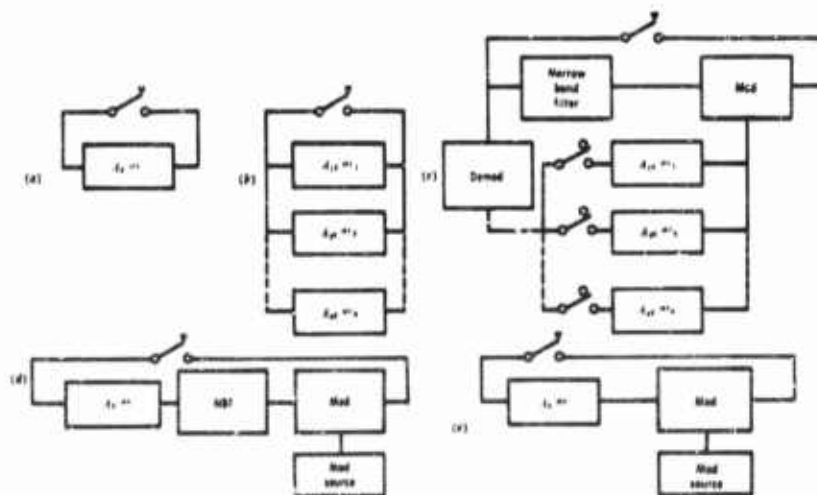
FIGURE 16-8   Examples of Repeater-like Jammers

example, the systems block-diagrammed in Figure 16-8. Figure 16-8a is an idealization of the simple time-shared repeater. In the system of Figure 16-8b, a number of simple repeaters are operated in parallel. A "babble of voices" jamming signal can be reradiated by using various values of delays (of the order of seconds or longer) in the several channels. In practice, such a jamming signal would usually be generated with a system analogous to Figure 16-8c. In this system, samples of the modulation signal are obtained and stored in a tape memory. In the meantime, the carrier of the received signal is obtained; a narrow-band filter is employed for this purpose in the illustration. A jamming signal is then obtained by modulating the extracted

*The duty factor for reception of some equipments discussed below may, however, be rather small.

carrier signal simultaneously with several of the stored modulation samples.

The system of Figure 16-8d permits modulation of the extracted carrier with waveforms having any desired dependence on the properties of the received signal. Figure 16-8d is in essence a spot jammer and therefore differs greatly from a simple repeater. On the other hand, the systems of Figures 16-8b and 16-8d may be essentially equivalent; for example, equivalent babble of voices jamming might be generated with the two circuits.

The systems of Figures 16-8c and 16-8d have a weakness inherent to spot jammers in that they must be tuned to the channel which is to be jammed. Figure 16-8e is a broadband counterpart of Figure 16-8d, but has the great disadvantage that the sidebands of the received signal are present in the retransmitted signal. One is again confronted with the danger that communication may actually be improved by the repeater. The effectiveness of such a jamming equipment is therefore critically dependent on the masking properties of the modulation introduced at the jammer.

Voice communication jammers analogous to the systems of Figures 16-8c, d, and e have been constructed and evaluation studies are continuing (References 6, 7, 8, and 9). Conclusions as to the relative effectiveness of the various schemes are particularly complicated when multiple-purpose equipment, such as a jammer intended for use against both AM and FM communications, are under study.

16.8 The Use of Repeaters Against Coded Radio Transmission Links

16.8.1 Introduction

It is convenient to classify coded radio transmission links as asynchronous although the delineation is not clear-cut. Examples of links which are regarded here as asynchronous include those which are hand-keyed, and common teletype. In the latter case, start and stop pulses are transmitted with each character, a character consisting of some small fixed number of pulses for a given coding. These start and stop pulses may actually be used to obtain synchronization within the individual characters, but such links are regarded as asynchronous in this delineation. By contrast, synchronism is maintained for periods of the order of days in other (synchronous) links.

Asynchronous coded radio links often employ frequency-shift keying. In any event, the information bandwidth of asynchronous transmissions is often small, the bandwidth of the spectrum of the transmitted signal is often comparable to the information bandwidth, and there are essentially a set of characteristic frequencies of transmission. For example, there are two characteristic frequencies in the case of a Simplex teletype channel. As would be expected, efficient jammers of asynchronous links such as those mentioned above radiate spectra concentrated near these characteristic frequencies.

In the case of common teletype, one type of error which can be produced by jamming and known as "hit" results from interference with the character-by-character synchronization (that is, by inducing errors in the time of character read-outs). An error in one or more characters may be generated in this way each time character synchronization is broken. It has been found that a major reduction in transmission efficiency can be obtained reliably, in the case of a Simplex channel, if a jamming to signal J/S ratio greater than about six decibels can be achieved at the receiver input. Such jamming can be accomplished with a variety of devices, which have output spectra concentrated near the characteristic frequencies of the target link, including repeater-like devices similar to those discussed in connection with voice communication jamming.

Equipment of special design can be employed for coded communications over distances short enough so that a single reliable path exists. Hits can be greatly reduced in this way at a fixed J/S ratio. For example, prediction circuits can be utilized to improve the probability of synchronizing on the proper pulse. In general, one finds that as the degree of synchronism of the target link is increased, the jamming problem can be made increasingly difficult.

Synchronization of a transmission link permits increased reliability of transmission in the presence of jamming (anti-jamming (A-J) advantage), or increased message security, or hiding, or some combination of the above, as compared with asynchronous links. These improved link properties can be obtained by transmitting a signal, some property of which is pseudo random. By this is meant that this property of the transmitted signal is being varied in a pattern related to the level variations of a (binary) pulse sequence. Short samples of these pulse sequences, which are often generated with a shift-register, look much like a random telegraph signal. However, these pulse trains are in fact periodic (commonly having periods of the order of days) and can readily be reproduced at a receiver for use in decoding provided the necessary interconnections of a shift-register generator have been made known (Reference 10). Correlation techniques or higher specialized demodulation schemes are employed in decoding at the receiver of such a link (Reference 11). In general, the decoding is strongly dependent on synchronization and on an exact knowledge of the pulse sequence employed in encoding at the transmitter. The requirement of brevity permits a discussion of only the following example of the many possible coding schemes which can be employed.

One encoding scheme (Reference 12) depends on the generation of a pair of binary sequences at a transmitter. One bit of information can be transmitted in some time interval through exercising a choice as to which of

the two pulse sequences is employed in constructing a signal for transmission in this time interval. The chosen pulse sequence might be passed through a narrow band filter, augmented with noise, and the resulting waveform used as a modulation signal during the transmission interval. At the receiver, a decision as to which of the two possible pulse trains was employed in generating the modulation waveform is made after determining the correlation between the received signal and signals generated in a similar manner at the receiver with each of the pulse sequences.

The autocorrelation function of the sequences discussed above can be made to approximate the sketch of Figure 16-9. The width of the spikes of the autocorrelation function can be made smaller by simply decreasing the pulsewidth, which produces an increase in the bandwidth of the spectrum. The A-J advantage of a synchronous
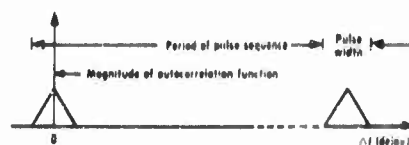
FIGURE 16-9 Idealized Autocorrelation Function of Binary Pulse Sequence

link employing pseudorandom encoding depends on the use of a correlation type receiver and on the use of a large number of pulses for each transmitted bit. The use of a large number of pulses having a pseudorandom relation for each transmitted bit produces a transmission spectrum which is very wide for a given rate of information transfer, and simultaneously makes it difficult for a jammer to achieve the requisite correlation between the jamming signal and the transmitted signal. An increase in message security, on the other hand, is more fundamentally related to the obtuse relation between the message and its encoded form (i.e., the encoding) rather than on the bandwidth of transmission. Hiding of a transmission is enhanced by obtaining a noise-like encoding, a broad transmission spectrum, and transmission at the lowest practical power level.

16.8.2 Jamming Schemes Depending on Multipath Distortion

Difficulties are encountered in operating either asynchronous or synchronous links over distances such that a single reliable path does not exist. Ionospheric transmission often occurs simultaneously over paths having a variety of lengths, and the propagation of the individual paths is often sporadic. Certain basic techniques which have been used to improve transmission with asynchronous links are referred to as space, frequency, and time diversity (Reference 13). These techniques are mentioned below. Certain jamming possibilities arising from the use of these techniques are also discussed. Attention is given specifically to the difficulty of multipath distortion in synchronous links and certain related jamming possibilities.

Space diversity is a technique for combating multipath distortion in which a number of receivers are disposed at various ranges from the transmitter. Space diversity is suggested by the fact that little correlation is observed among the fading of the signals obtained at several receivers, provided they are sufficiently separated. In frequency diversity, on the other hand, transmission occurs at a multiplicity of frequencies. Again, the merit depends on an essential independence of fading properties for widely separated frequencies. Finally, time diversity systems depend on obtaining received signals which are distributed in time.

Systems employing any of the above techniques, or combinations of them, provide a number of signal samples from which one wishes to make a reliable estimate of the transmitted signal. The processes of detection, weighting, and combining are used in forming an output signal. These processes occur in various orders in extant systems, and a number of weighting schemes have been employed. For example, one may simply choose the signal sample having the highest average power and ignore the remainder. Uniform weighting of all signal samples has also been considered, as well as weighting according to the average power of the signal samples. Each of these various weighting schemes has advantages in certain applications.

The potential of a repeater-like jammer against asynchronous communication links employing diversity techniques is dependent on the weighting scheme used. When the weighting depends on received signal pow... it may be possible to build up the value of the weighting function for one signal with strong transmissions from a repeater. Repeater transmissions would then be interrupted and noise jamming could ensue for the persistence period of the weighting circuits. Such a jammer would have no special merit against a link employing fixed weightings of the several received signals.

Space diversity is often an impractical technique for solving the problem of multipath distortion in synchronous pseudorandom links. This arises from the fact that the spikes of the autocorrelation function in such a link are usually very narrow. As a result, all the paths of effective received signal components must be very nearly of the same length. Typically, only a few such paths will exist. This is to be contrasted with links utilizing more efficient coding. These links may have a larger number of effective paths as a consequence of their wider autocorrelation spikes. One would expect the transmission properties to become independent more rapidly, as the sites are separated, when they are derived from a large number of paths of widely distributed length than when they are derived from a few paths of nearly the same length. Two specific schemes for combating multipath distortion which have been used in pseudorandom links and which are essentially time diversity techniques, are mentioned below.

One scheme (τ-tracking) (Reference 12) for reducing multipath distortion in pseudorandom links involves the use of three correlators at a receiver. Two of these correlators are operated at delays differing by approximately the width of the autocorrelation spike of the transmitted signal. A variable delay is adjusted continuously to hold an error signal, made up of the difference of the two correlator outputs, near zero. The output is taken from a third correlator operating at a delay value intermediate to the delays of the correlators in the control circuit. The time constant of the correlator delay control circuit is such as to permit the receiver to track the typical slow variation of effective path length of the link, i.e., tracking in delay τ.

It has been proposed that a repeater-like equipment in a favorable tactical situation could lower the transmission efficiency of a link employing τ-tracking by "range-stealing and dumping." In this scheme, a repeater induces some delay setting at the correlator with strong transmissions. The correlator is then "dumped" at this delay by interruption of repeater transmissions. The jammer may even transmit noise intermediate to repeating. In order for this scheme to be successful, the delay of the repeater transmission path must fall within the range of delay times in which normal link operation is permitted. The repeater must therefore be located more or less between the transmitter and repeater, the precise requirements depending on the variability of the receiver correlator delay. Such a jammer site may not be obtainable.

RAKE (Reference 14) is a scheme for combating multipath distortion in which a delay line with a large number of fixed taps is employed at the receiver. Demodulation correlators are employed at each tap. The receiver output is made up from the weighted outputs of the several correlators. The correlator outputs are weighted according to their average output powers over some relatively long period. Repeater-like devices appear to be one of the most promising jammer types for use against RAKE. The weighting distribution on the outputs of the correlators can in principle be modified by reinforcing transmission strongly with a favorably disposed repeater. The receiver channels which are "opened up" in this way are then subjected to noise jamming for some time interval. Again, the availability of a favorable jammer site is critical.

16.9 Some General Remarks Concerning Jammer Look-Through Capability

Jammers with provisions for monitoring while jamming (look-through capability) have been sought widely. Information obtained using look-through may permit an increase in jammer effectiveness by:

1) increasing the jamming duty factor
2) improving set-on accuracy
3) confining jamming to periods of target activity
4) adjusting jamming signal parameters continuously with variation of target parameters.

The 100 percent duty factor (two-antenna, repeater, which has not been successfully employed for fuze or communications countermeasures, is an example of a jammer with look-through capability.

Construction of a communications jammer with look-through capability, referred to as CMJS (continuous monitor jamming system), was completed by the Electronic Defense Group of The University of Michigan in 1957 (Reference 15). This equipment radiates a *continuous* FM jamming signal. Leakage from the transmitting antenna to the receiving antenna is utilized as a local oscillator signal at the look-through receiver of this equipment. A target signal is indicated when an output is obtained from a special low-pass i-f section in the receiver. Both AM and FM modulations on the target signal may also be reconstructed (read-through capability), using a correlated sampling of the output of the receiver i-f section.

Monitoring of target signals can also be accomplished by time-sharing between reception and transmission. Examples of such equipment include time-shared repeaters and the EDL "Hunt and Lock-On System", (Reference 16), which has proven effective against simple CW fuzes. The Hunt and Lock-on System operates in a search mode until a target signal is selected. It then operates as a spot jammer for some predetermined interval by radiating an internally generated carrier which is square-wave modulated at an audio rate. Set-on is maintained by monitoring of the target signal during the transmitter dead times. When the set-on accuracy is maintained satisfactorily (to within perhaps 5 kc for typical small CW fuzes), jamming effectiveness against a single fuze is comparable to that of the long-delay repeater. The Hunt and Lock-On System has a very narrow instantaneous bandwidth and usually processes target fuzes individually. The time rate at which fuzes can be processed with this equipment is therefore limited. By contrast, extant repeaters can handle large numbers of fuzes by power-sharing among them at a commensurately reduced range. Monitoring schemes which depend on time-sharing endanger jammer effectiveness when the length of the reception periods is comparable to the reciprocal of the bandwidth of the target equipments.

REFERENCES

1. Barton, B. F., "A Modification Study of the AN/MLQ-8(XE-3)," University of Michigan Research Institute, Technical Report No. 91. Electronic Defense Group, 2899-3-T, May 1959. (SECRET)

2. Miller, M. H., and J. A. Boyd, "Basic Principles of Doppler Fuze Countermeasures," University of Michigan Research Institute, Part I, Technical Report No. 4, Electronic Defense Group, August 1952. (SECRET)

3. Beattie, L. A., J. L. Deutermont, R. J. Hansen, F. E. Pickel, and P. H. Rogers, "Design of the AN/MLQ-8 Repeater Jammer," University of Michigan Research Institute, Technical Report No. 55, Electronic Defense Group, 2262-101-T, January 1956. (SECRET)

4. Schneiderman, S., S. Lenefsky, and H. Schajn, "Interim Engineering Report on Design, Development and Construction of Countermeasures Set Electronic AN/MLQ-8(XE-3). A Multichannel Repeater Jammer," Report No. 341-Q 15, Instruments for Industry, Inc., Hicksville, N.Y., December 1958. (CONFIDENTIAL)

5. Morton, C. W., "Results of Field Tests of 'Blue Warrior'," Technical Report No. M-1833, U.S. Army Signal Engineering Laboratories, Evans Signal Laboratory, March 1957. (SECRET)

6. Rothschild, D. R. and D. H. DeVries, "Effectiveness of Jamming Signals Against an AM Voice Communication Receiver," University of Michigan Research Institute, Technical Memorandum No. 51, Electronic Defense Group, 2262-172-T, November 1957. (SECRET)

7. Hellwarth, G. A., D. E. Ross, and D. R. Rothschild, "Jamming of FM Voice-Communication Receivers, Including the Effect of Receiver Detuning," University of Michigan Research Institute, Technical Report No. 95, Electronic Defense Group, 2899-7-T, January 1960. (SECRET)

8. Hellwarth, G. A., D. Ross, and D. Rothschild, "Discussion of Sampled-Speech Phenomena Occurring During Jamming of Voice-Communications Receivers," University of Michigan Research Institute, Technical Memorandum No. 63, Electronic Defense Group, 2252-195-T, October 1958. (SECRET)

9. Hellwarth, G. A. and D. E. Ross, "Evaluation of Time-Shared Jamming Signals Directed Against Voice-Communication Receivers," University of Michigan Research Institute, Technical Report No. 108, Electronic Defense Group, 2899-27-T, June 1960. (SECRET)

10. Birdsall, T. G., and M. P. Ristenbatt, "Introduction to Linear Shift-Register Generated Sequences," University of Michigan Research Institute, Technical Report No. 90, Electronic Defense Group, 2262-189-T, October 1958. (UNCLASSIFIED)

11. Green, P. E., The Output Signal-to-Noise Ratio of Correlation Detectors, *Trans. IRE*, PGIT, IT-3, 1957.

12. Green, P. E., "The Lincoln F9-C Radio Teletype System," Massachusetts Institute of Technology, Technical Memorandum No. 61 & Addendum, Lincoln Laboratory, May 1954. (SECRET)

13. Bond, F. E. and H. F. Meyer, The Effect of Fading on Communication Circuits Subject to Interference, *Proc. IRE*, Vol. 1, No. 5, pp. 636-642, May 1957.

14. Price, R. and P. E. Green, A Communication Technique for Multipath Channels, *Proc IRE*, Vol. 46, pp. 555-570, March 1958.

15. Orr, L. W., "Secure Communications Systems Using the CMJS," University of Michigan Research Institute, Technical Report No. 78, Electronic Defense Group, 2262-176-T, February 1958. (SECRET)

16. "Evaluation of ECM Hunt-Lock-on Technique," Final Report USAEPG-SIG, 920-133, U.S. Army Electronic Proving Ground, Fort Huachuca, Arizona, July 1958. (SECRET)

17

Programmed Automatic Jamming Systems

K. H. BARNEY

17.1 Introduction

Automatic operation of equipment becomes a necessity when lack of time does not permit a human to make a quick decision. Electronic-warfare equipment can be operated with momentary notice if programmed controls are used. The use of programming means that decisions have been made concerning how and when the equipment should operate under expected conditions. Programmed systems may be simple or complex. For example, a simple programmed system might consist solely of a repeater designed to function against proximity fuzes of a specific type; in this case, a decision to exclude other fuzes has been made before the repeater was designed and constructed. A complex system could be one such as the airborne programmed jamming system discussed in this chapter. This example is given only to indicate the types of problems encountered when designing such a system.

Specifically the typical types of processing and decision requirements necessary in an airborne programmed jamming system are discussed. The components which may be found in a jamming system, e.g., jammers, intercept receivers, antennas, are described in other chapters. Considerations in the use of these equipments and the role of airborne jamming in the protection of aircraft are also discussed in other chapters.

17.2 Processing and Decision Requirements

17.2.1 General

This section explores the brain of the system—the programmer—and deals with the major characteristics of the "brain" and the types of equipment which may be used for this function.

It should always be remembered that the programmer is merely a means of automatically implementing the wishes of its designers and users; like all computers, it will never make decisions any more wisely or any more accurately than is warranted by the accuracy of the information fed into it during or before the flight. Its chief advantage is that it can assimilate large quantities of data and react to them far faster than a human operator, although its "judgment" can never approach the judgment of a human being. Another point that has been raised before is that the programmer does not have the same sense of responsibility for the safety of the aircraft as a human operator and cannot be trained through experience to do better, nor can it be courtmartialed when proved incompetent.

Inputs to the programmer are derived from two sources: preflight instructions read into the programmer and intercept information received during the mission itself. The preflight instructions basically establish what the jammer should do when various combinations of input signals are received. These instructions may be of the type that established the priority of jamming modes where the jamming equipment is capable of generating many different types of jamming signals. These priorities may be established as a function of geography, of whether the aircraft is in area-defense or local-defense regions, of whether the aircraft is alone or part of a mass raid, and of whether or not other electronic equipment incompatible with the jammer is in operation. In-flight information to the programmer may consist of all or part of the interceptor receiver data related to the various characteristics of the received radar signals or simply the receiver's evaluation of these signals and subsequent classification of them into radar types, such as early warning, GCI, acquisition phase of tracking radar, tracking radar, and seeking missile. Other in-flight information may be related to the aircraft, such as aircraft location, altitude, and velocity. Since many degrees of complexity are possible, the requirements for the programmer must be carefully related to the capabilities of the transmitter and the reaction time desired.

17.2.2 Elements of Programmer Control

Typical elements of programmer control are shown in Figure 17-1. Inclusion of all or part of those functions in a specific programmed automatic jamming system is, of course, subject to a variety of considerations. However, in order to aid the equipment designer to establish his own design requirements, the possibilities of each element of programmer control will be discussed in some detail. An example has been chosen in which a number of jamming equipments, each covering a different frequency band and each capable of operating in several distinct modes, are to be controlled in one or more aircraft installations.
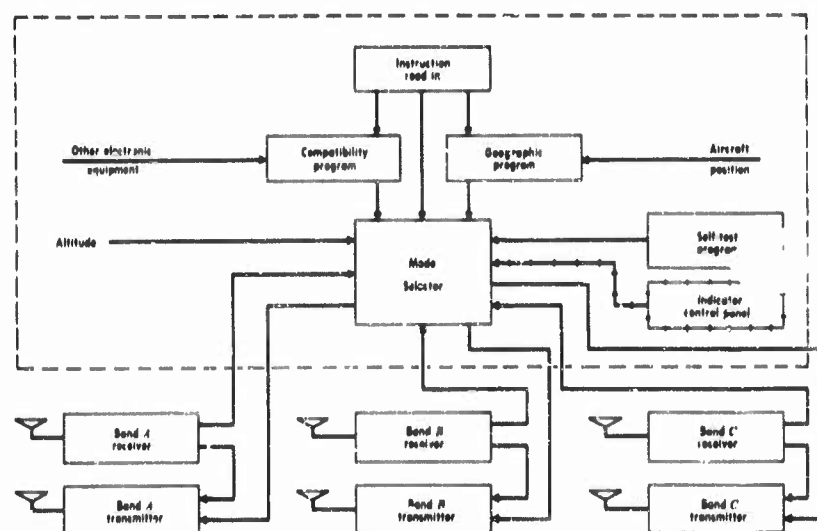
FIGURE 17-1 Elements of programmer control.

17.2.3 Mode Selecting

The mode selector may be considered as a group of switches and switching circuitry which activates the jamming transmitters as a result of data intercepted by the receiver. The manner in which the switching circuitry is set up and the numbers and types of combinations possible are a function of the instructions fed to it by the compatibility and geographic programmers, the preflight instructions, and the control panel (if used). Examples of types of input data fed to the mode selector from the receivers are:

(a) Frequency
   Coarse—by band
   Fine—to the ultimate precision of the intercept receiver

(b) Type of Radar Intercepted
   Search, track seeker missile, etc.
   Signal from other jammer

(c) Number of Signals by Band
   Density of radar signals

(d) Passing of Main Lobe of Radar

(e) Modulation of Radar Signal

Typical decisions made by the mode selector are:

(a) Proper mode of operation, e.g., barrage noise, multispot noise, false-target deception.
(b) Sequence in which modes are to be initiated.
(c) Look-through cycle, if required.
(d) Modulation to be applied to jamming signal.

It is obvious that the capacity of the mode selector depends upon the following:

(a) Total number of input signals and input signal levels measured.
(b) Total number of individual instructions.
(c) Total number of individual decisions to be made (control busses to be activated).

Not so obvious is the effect of saturation by the radar environment upon the receiver-programmer combination. Saturation of these elements may come about through the interception of a large number of radar signals at different frequencies in the jammer band or by the victim's use of a high power, high-duty factor radar, possibly as a decoy, to influence adversely the mode of operation of the jammer. An example of this would be the use of a high-power tracking radar in an area-defense environment. In this case, the jammer might not respond properly to GCI radars if its look-through cycle and output power level were selected by the mode selector to favor the tracking type radar in response to signals received from the decoy radar(s). This situation may be avoided if there is sufficient capacity in the receiver-programmer to recognize a large number of individual signals simultaneously received so that decisions may be made on the basis of all the input signals rather than that of a single overriding signal. Saturation may also be avoided by instructing the receiver to reject high-duty factor signals or (if sufficient intelligence information was previously available) by instructing the receiver not to look at frequencies used by these radars. Saturation of the transmitter may be avoided when pulsed countermeasures are employed by limiting the duty factor of the transmission to an acceptable, predetermined amount. Finally, the effects of saturation may be reduced by programming a change of mode of jamming to one that is basically less saturable, for example, noise barrage (wide or narrowband) jamming.

The mode selector may be required to assign the look-through period (if one is employed) depending upon the probable radar type. The implementation of this signal will depend upon whether range or angle jamming is required, or both. These considerations apply equally to search-type or track-type radars. When, for example, range jamming of the main beam of a search radar is desired, a low duty factor jamming may be used if jamming can be initiated just prior to the passing of the main beam and stopped

shortly thereafter. In this case the system will be in a receive condition for a major portio.; of the time. On the other hand, the range jamming of a radar which is constantly illuminating the jammer aircraft requires a high duty factor type jamming which is interrupted only rarely by receiver look-through periods.

The preceding paragraphs treated examples of the types of decisions that the mode selector of a programmed automatic jammer is required to make. Before the design of the mode selector is initiated, it is wise to prepare a table of inputs *versus* responses in order that the capacity required of the mode selector may be determined. This table should indicate each input signal, the precision to which it is measured, and the resulting commands which can be programmed.

The degree of possible mode selection flexibility is directly related to the capacity of the mode selector. There are various approaches to the design of the mode-selector equipment depending upon the flexibility of programming required. These are shown below.

Type	Response Time	Flexibility of Programming
Fixed-wired	Short	Nonflexible
Variable-wired (plug in wired program punched card or tape)	Long	Flexible
High-speed memory (magnetic drum or tape)	Short	Flexible

The fixed-wired type is useful only when the automatic program need never be changed. In this case, the equipment is permanently wired so that certain combinations of input signals always produce the same jammer outputs. This approach cannot be used where it is necessary to alter any programming requirements because of changing missions or changing tactics during a mission. It cannot be used in any but the simplest jammers.

The variable-wired type makes use of programming changes achieved by plug-in wiring boards which can be changed on a preflight basis. If program changes are required during the flight, these instructions can be stored on punched cards or punched tapes. This approach provides a great deal of operational flexibility and can be used when the number of separate decisions and levels of measurement of input data is fairly restricted.

When extreme operational flexibility is required and when the number of jamming modes and levels of measurement of input data is large, the use of a high-speed large-capacity memory such as a magnetic drum and its as-

sociated circuitry is indicated. A magnetic-drum type memory using conventional digital-computer techniques can typically provide the capability for making several hundred separate output decisions based upon 100 to 500 separate inputs in accordance with five separate programs for penalties on the order of 75 pounds, 3 cubic feet, and 50 watts of primary power.

17.2.4 Geographic Programming

It is often desirable to be able to change many of the logic equations as a function of geography or to meet changing tactical situations during the mission of the jammer aircraft. Such changes can be made by a geographic program unit which receives positional data from the navigational system of the aircraft. Examples of tactical zones which may have differing jammer requirements are as follows:

(a) Friendly zone
(b) Early warning zone
(c) Area defense zone
(d) Local defense zone

Geographic programming of the automatic jammer is possible only if the approximate geographical positions of these zones can be predicted in advance. Also important from a programming standpoint is the fact that the jamming requirements in each of zones b, c, and d may differ because of the raid tactics, which affect each aircraft's defense. These tactics fall into one of the following categories: (a) mutual support—mass-raid tactics; (b) self protection—single-aircraft tactics.

Some examples of geographic programming will illustrate the flexibility that it lends to the programmed automatic jamming system and the operational needs that it fulfills. It is assumed that the aircraft or unmanned vehicle is on a strategic bombing mission.

The mode of operation of the jammer in the friendly zone a would probably be the standby mode or at most the receive-only mode. Upon reaching the early warning zone, the geographic programmer would enable the mode selector to select only the modes of operation that are effective against early warning radars and would permit jamming to be initiated only after a preset geographic check point had been passed. (For purposes of over-all raid coordination or synchronization, a preset time to initiate jamming might be utilized in place of the passing of a preset geographic check point.) The types of jamming used in the early warning phase might be aimed at deceiving the victim into expecting a large attacking force to arrive from one angle, whereas in reality it is planning to attack from another. This tactic might succeed in diverting the defending activity from the main striking force. Other modes of jamming useful in this phase might be those which

prevent the enemy from correctly assessing the strength of the striking force. Noise jamming—barrage or multispot—and multiple false-target deception would be useful modes for these types of jamming operation. In the case of noise jamming the mutual-support concept would involve many aircraft utilizing noi e-jamming modes of operation simultaneously. In the case of multiple false-target deception the mutual protection for many aircraft could probably be generated satisfactorily by only one jammer operating in this mode. The jammers in the other nearby aircraft could remain silent at this time or operate in noise-jamming modes, thereby creating a combination of multiple false-target deception and noise jamming on the presentations of the enemy early warning radars. The possibility of complete jammer silence during the early warning phase should not be overlooked in that it may be advantageous to try to slip through a weak point in the enemy early warning net undetected. The use of any jamming by aircraft attempting this tactic would, of course, defeat the purpose of the maneuver.

When the aircraft reach the area-defense zone, the modes of jamming described above may still be effective in preventing the vectoring of enemy interceptors to the bomber's position and the assessment of raid strength. In addition, it may be desirable to program the mode selector not to respond to tracking radar signals if the jamming of search-type radars were degraded by this action. The decision to program the jammer in this fashion would have to be made taking into account at least two factors: (a) whether or not the jammer was providing protection for other aircraft in this zone, and (b) the effectiveness of the weapons in the area-defense zone utilizing tracking-type radars *versus* the effectiveness of the weapons employing the search radars.

On the other hand, in the local-defense zone the programming of the mode selector may be such as to favor the jamming of tracking radars to the complete exclusion of search-type radars. Modulated noise, inverse modulating repeaters, and phase-front warping types of jamming might be most effectively programmed in this zone.

Physically, the geographic programmer may consist of a punched tape or punched-card-type memory unit. One section of tape or group of cards would be punched with the instructions for the mode selector applicable to a particular geographic zone. Separate sections of tape or groups of cards would contain the instructions applicable in the other geographic zones. Passage of a preset geographic check point designated by small increments of latitude and longitude surrounding the nominal latitude and longitude of the check point would cause the appropriate section of the tape or groups of cards to instruct the mode selector in the modes of jamming permissible for that geographic zone. These check points would be established prior to the mis-

sion based on the battle plan of the particular aircraft and would be read into the geographic programming unit on a preflight basis.

A further use of geographic programming is to minimize incompatible operation between the jammers themselves or the jammers and other friendly electronic equipment. Several aspects of the compatibility problem are discussed in the following section.

17.2.5 Interference and Electronic Compatibility

Electronic compatibility problems arise whenever equipments that receive and transmit are used in proximity to each other. The receive and transmit functions do not necessarily have to be automatic for a compatibility problem to exist. The problem is aggravated when all the equipments in question receive as well as transmit. The problem is further accentuated when one of the equipments is designed to jam many of the equipments of the same type with which it is required to be compatible.

The compatibility problem is broken down into the following categories:

(a) jammer to own aircraft equipment (intra-aircraft)
  (1) Other jamming equipment
  (2) Other electronic equipment
(b) Jammer to other aircraft equipment (inter-aircraft)
  (1) Jamming equipment
  (2) Other electronic equipment

Interference between the electronic equipment listed above can occur in a number of ways. These include response of the receivers (fundamental or spurious response) to the fundamental or spurious response of the transmitters, and responses of the receivers to the modulation products of several transmitter frequencies generated in the receiver's rf stages or mixer.

There are no simple, all-purpose solutions to the compatibility problem. However, attacks on the problem can be made by the use of the following techniques separately or in combination:

(a) Isolation of antennas and equipment shielding.
(b) Filtering of transmitter outputs and receiver inputs.
(c) Use of compatible modes of equipment operation.
(d) Time sharing of equipment operation.

The first of these items is governed strictly by the installation of the equipment in the intra-aircraft case and by the spacing of the aircraft in the inter-aircraft case. The second technique may be effectively used if the insertion losses of the filters permit acceptable reductions of transmitter output power or receiver sensitivity, and if the gaps in the jammer spectra (receive or transmit) can be tolerated. The remaining techniques can be

employed through the proper programming of the system to lessen many compatibility problems.

It is obvious that answers to the compatibility problem must be stated in terms of the actual equipments involved. However, a few examples of approaches are given which may indicate a means to the solution of a particular problem. The discussion is confined to techniques c and d as applied to the jammer-to-other electronic compatibility problems since these involve the programming of the jamming system.

Interestingly enough there may be some modes of jammer operation which are compatible with certain electronic equipment. If this is the case, the mode selector can be arranged such that, whenever a particular nonjammer equipment must be used, the jammer is programmed into a compatible mode of operation. One example of this type of operation involves the bombing and navigation radar. At least two frequently used modes of jammer operation are likely to be wholly or partially compatible with this radar. These are pulsed track-breaking repeating deception and mutiple false-target deception. These modes provide countermeasures effective against tracking radars and search radars, respectively. The repeater, such as that described in Chapter 15, will have no effect upon the bombing and navigational type radar since the repeated pulse will occur during the radar main bang (intra-aircraft case) or in coincidence with the radar echo (inter-aircraft case). The only effect of the radar upon the repeater may be degradation in its inverse modulation capabilities if the repeater is capable of handling only a few radar signals at a time and a degradation in its total signal-handling capability if the repeater transmitter is duty-factor limited. These latter effects may be entirely eliminated in the intra-aircraft case by sending a blanking pulse to the repeater that is coincident with the radar main bang. This is a time-sharing technique that is discussed to a greater extent below.

The effect of multiple false-target deception on the bombing and navigation radar is to put a pattern of blips on its presentation which may probably be distinguished from land targets by the operator of the radar. This situation may be satisfactory in all but the worst cases since it is likely that the character of false target blips is entirely different from the ground-return patterns to be interpreted by the operator. This interference could, of course, be removed in the intra-aircraft case by blanking the radar with pulses coincident with the false-target deciption pulses transmitted. The effect of the radar upon the jammer in this case is merely to use some of its output power capability. If this is intolerable, the jammer may be blanked in the intra-aircraft case by pulses coincident with the radar main bang.

The preceding paragraphs give examples of compatible modes and also indicate possible use of time sharing in the form of blanking pulses. Where

no compatible modes exist, time sharing can also be employed to effect compatibility to a certain degree. Consider the case of noise jamming and the operation of the same bombing and navigation radar. Normally the noise jammer and the bombing and navigation radar would not be compatible in the intra-aircraft case (assuming insufficient isolation and/or filtering). However, two types of time sharing could be employed to achieve compatibility. The first means would be to use the bombing and navigation radar intermittently for short intervals (perhaps during the noise jammer look-through periods) and to transmit noise the rest of the time. It has been estimated that operation of a typical bombing and navigation radar on a $\frac{1}{10}$ duty-factor basis may be satisfactory for most bombing runs. Thus, this type of operation may not excessively degrade either the jammer or the radar effectiveness.

Another time-sharing technique which might increase the jamming effectiveness by increasing the jamming duty factor (above the $\frac{1}{10}$ duty factor of the preceding example) is to employ a wide range gate synchronized with the radar to blank the noise transmission. Thus the radar could "see" without any interference over a narrow range increment whose range position with respect to the aircraft could be readily adjusted by the radar operator. Of course, this technique is effective only in the intra-aircraft case.

Time-sharing compatibility in the inter-aircraft case can be achieved by transmitting synchronizing signals on an inter-aircraft communication link, by coding the basic transmissions by geographic programming, or by time-synchronizing the programs of the jamming systems. The latter technique would employ very accurately synchronized clocks in each aircraft.

17.2.6 Self-Test Program

An additional feature of programmed automatic systems that is desirable if the weight, volume, and primary power allocated is sufficient is a means of self-checking the programmer. Self-checking techniques should be used only to increase the reliability of the jamming system. Whether or not they are used will largely depend upon the complexity of the programmer and the number of jamming modes available. Two major types of self-checking are (1) component-test and (2) logic equation loop input and output check. The first involves periodically programming the mode selector (or the other components) with a test program to determine if all the components are functional. The second involves a continuous check on the programmed jammer to determine if the jammer is actually functioning in the mode programmed. If it is not, another jammer mode of operation is automatically selected.

18

Confusion Reflectors

A. T. GOBLE

18.1 Introduction

The use of devices to produce radar echoes other than those of the proper radar targets was proposed in the early days of the development of radar. Some of the early history and some of the general principles of the use of such confusion reflectors have already been mentioned in previous chapters. Such reflection devices naturally find a wide variety of applications difficult to separate out into isolated chapters. Material to be presented in this chapter will be an essential part of the next two chapters; material to be given in them could have been included here but has been deferred to them.

In the context of electronic countermeasures, the function of any reflecting device basically is to introduce echoes which divert attention from the proper targets of any radar. In doing this, a form of noise is introduced into the radar system. The seriousness or effectiveness of this noise is likely to depend upon many circumstances and is usually difficult to evaluate. If the proper target is completely immersed in a field of many equivalent false targets, so that no identification is possible, the effectiveness is complete. This is seldom the case because the echoes from the confusion reflectors are rarely equivalent to those from the proper targets, but in every case there will be some degradation of the radar system by the mere requirement that the extra echoes must be inspected and rejected. The effectiveness of the countermeasure is clearly greater when the identification of the proper target from the false targets is made more difficult. Mere quantity of confusion reflectors will

often increase this difficulty. Any other technique which improves the simula-
tion of the proper target increases the effectiveness of the noise. For example,
confusion reflectors are likely to differ from most proper targets in their lack
of motion. If equivalent mobility is given to the reflector, it increases its ef-
fectiveness. This is one way in which a "mere" confusion reflector can be
turned into a decoy. Further discussions of such subtleties are postponed to
Chapter 20.

Customary thinking divides the uses of confusion reflectors into (1) screen-
ing against tracking radars and (2) confusion of surveillance systems. Al-
though both applications have elements in common, they are relevant to
different parts of the defense system. Confusion reflectors *cannot* provide a
screen that hides a target beyond it. They may hide a target *immersed* in a
cloud of them and they may degrade the surveillance system by the clutter
they produce.

Nearly all applications of confusion reflectors are such that the efficiency
of the reflector is improved by decreasing the packaged size and weight for
the same radar scattering cross section. Three essentially different systems
are used to obtain large cross sections for small size and weight. These are:

(a) Resonant dipoles—Chaff
(b) Nonresonant streamers—Rope
(c) Corners or other reflective geometries—Angels

The code names following each are those generally used for the airborne
versions. The more general term used during World War II was *window*;
but the term *chaff* has now nearly replaced this in the language of the United
States Armed Forces. Angels have found little application because of the
difficulty in building units that are sufficiently rigid for dispensing from
rapidly moving aircraft. They have application in decoys and target modifi-
cation schemes.

One of the favorable features of confusion reflectors is that they may be
designed to have wideband-frequency characteristics. For example, a single
package capable of imitating standard aircraft over a range of frequencies
from 50 to 10,000 megacycles (and probably still higher) is now available.
It may not be the most economical package to use, but it illustrates the pos-
sibilities of wide frequency coverage.

Another feature is their essential simplicity. In fact, the most difficult
problem in their use by aircraft is providing a suitable dispensing system
along with the penalties imposed because of the additional weight to be car-
ried. Although a solution has been found for most aircraft, a few aircraft
systems seem to be quite difficult to adapt to suitable chaff dispensing. They
may not have any space available, or the location of the space may not favor
chaff dispensing. It usually seems to be desirable to dispense into a region

of high turbulence reasonably forward. A compartively simple and reliable mechanism is now available, and standard approved installations have been developed for most aircraft. Little servicing is required. The actual labor of loading the dispensers is not inconsiderable, but still reasonable.

Although the use of chaff has definite limitations and although a great deal of effort has been put into chaff countercountermeasures, it seems likely that as long as pulsed radar systems play a significant role in general defense systems chaff will continue to be very useful.

18.2 Theory of Response of Confusion Reflectors

The return to a radar that results from the scattering of a radar signal by a target is most easily indicated by giving its back-scattering cross section σ, defined by

$$\sigma = 4\pi \frac{\text{Power reflected back per steradian}}{\text{Power incident upon the target per unit area}} \qquad (18-1)$$

The power per unit area incident upon the target is proportional to the radar power, and the gain of the radar antenna and is inversely proportional to the square of the range. The response of the radar receiver will be proportional to the power reflected back per steradian to the antenna gain and receiver sensitivity and will also be inversely proportional to the square of the range. It will be seen that σ depends only upon the characteristics of the target. Another way of describing σ is that it is an area such that, if all the energy falling on that area were reradiated isotropically, then the intensity of the scattered radiation would be the same as that from the target.

The calculations of the back-scattering cross section for dipoles and non-resonant streamers are somewhat similar and will be considered before the calculation for the reflective devices such as corners. All theoretical work in this field has been aimed at finding the best approximation to the solution of Maxwell's field equations for the situation appropriate to chaff or rope. The incident electric field is taken to be known. Then an attempt is made to determine what currents in the conductor of the scattering elements are produced by this field. Once these currents are known, the field back at the antenna can be calculated. The real problem is to calculate the currents in the conductor since the reradiated field must be considered as much a part of the field as the incident part.

18.2.1 Chaff

For the case of chaff, several theories have been advanced by American authors. Undoubtedly other work has been done, but the author does not

have access to any other calculations. In all of these, the scattering chaff is considered to consist of a cloud of wire-like dipoles, all of the same length. The following additional assumptions are also made:

(a) The dipoles are separated so far that they do not interact.

(b) The dipoles are placed in a somewhat random spacing so that the phases from each dipole are random.

(c) The orientation of the dipoles is specified (usually random).

(d) The wires are perfect conductors.

The first calculation of this sort was made by Chu (Reference 1). He assumed that the current was distributed along the wire in a sinusoidal way. This implies that resonance occurs when the length is an integral number of half-wavelengths. His results gave the value $\sigma/\lambda^2 = 0.15$ for a half-wave dipole. This value is independent of the ratio L/a, where L is the length and a is the radius of the dipole. The dimensionless quantity σ/λ^2 is the convenient one to use along with another dimensionless quantity, L/λ, which is the principal independent variable.

A more realistic approximation is the one used by Van Vleck, Bloch, and Hammermesh (Reference 2). In their treatment, the tangential component of the total electric field (incident plus scattered field) for a single dipole is made zero. This leads to an integral equation for the current distribution which is then solved by approximation methods. The values for the current distribution so calculated can then be used to determine the value of σ. The results are dependent on the orientation of the dipole relative to the polarization and propagation directions of the incident wave. The results are averaged for all orientations. The results also depend upon L/a. The numerical results have been obtained for a wide range of values of L/λ^2 and of L/a. In the case of a flat strip (more appropriate to most chaff), one should replace a by $W/4$, where W is the width of the strip.

Figure 18-1 is a plot of some of these results. The full set of results is tabulated in a Standard Rolling Mills report (Reference 3). The following points should be noted:

(a) The resonant length does not come directly at $\lambda/2$, but instead at a slightly shorter length that varies with the length-to-width ratio, L/W.

(b) Although σ/λ^2 increases with W it does so rather slowly; so it is more efficient to use dipoles as narrow as possible, provided one does not go to such extremes that the resistance increases greatly.

(c) The bandwidth of the response also increases with W. This increase is small; so, here too, it is better to use dipoles as narrow as possible and to obtain increased bandwidth by using varying lengths.

Figure 18-1 does not indicate directly the behavior of chaff for a given
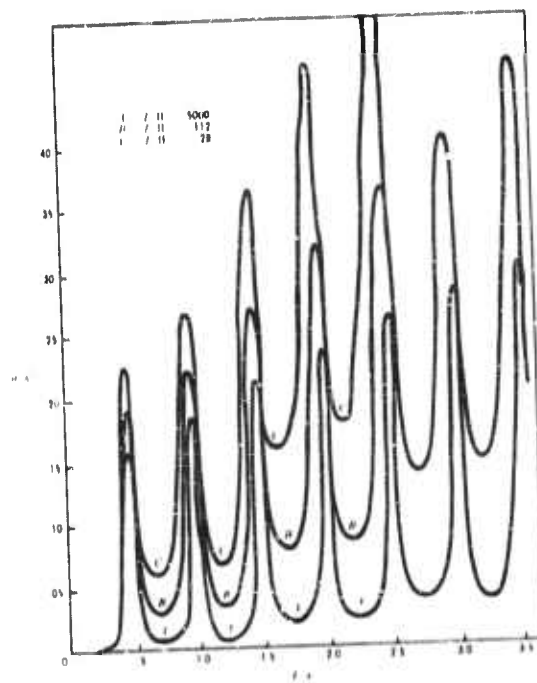
FIGURE 18-1. Normalized cross section—Van Vleck theory.
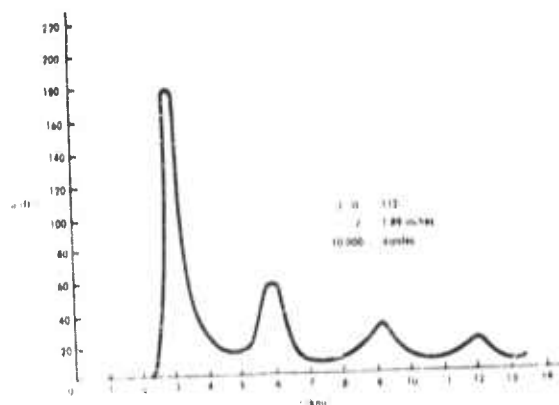
FIGURE 18-2. Dipole response versus frequency.

group of dipoles all of the same length as the frequency is changed. This is shown in Figure 18-2, where the theoretical echo versus frequency for a unit of dipoles all of the same length is plotted. It will be noted that at higher harmonics there is still some response. There are even important contributions in the nonresonant portions of the curve.

Recently, a new method for computing scattering cross sections was developed by Schwinger and Levine (Reference 4). This variational method has been used for a number of special cases (Reference 5, 6, 7, 8, and 9), but there is no indication that it will lead to any significant differences from the Van Vleck theory presented here.

18.2.2 Rope

The scattering cross section of nonresonant streamers, rope, depends very much upon the shape of the streamer as well as upon the wavelength of the radar. The actual configuration of a dispensed streamer cannot be predicted. The only theory published so far (Reference 10) assumes that the full streamer can be considered to consist of parts each of which is a helix with a large radius compared to a wavelength. The cross section according to this theory is

$$\sigma = \frac{KL\lambda/2}{\left[\ln \frac{\lambda}{\pi\gamma a}\right]^2 + \left[\frac{\pi}{2}\right]^2} \tag{18-2}$$

Here K depends upon the form of the streamer configuration, a is the equivalent radius of the streamer ($a = W/4$), λ is the wavelength, and $\gamma = 1.781$.

Figure 18-3 shows a plot of the predicted cross section for three 750-foot rolls of ¼-inch rope. There may also be a small effect when $W = \lambda/2$.

Long dipoles are difficult to store and to disperse. Except at low frequencies, rope is somewhat inefficient. To bridge the gap, it has been proposed to consider tied dipoles, a colinear array, called tuned rope. If the array is truly colinear so that coherent scattering occurs, the gain in cross section that is obtained in a narrow region perpendicular to the length leads to a loss in other directions. The usefulness of this scheme thus depends on the array not being colinear. The interaction between neighboring dipoles also tends to decrease the echo. This effect decreases the efficiency to such an extent that it is not
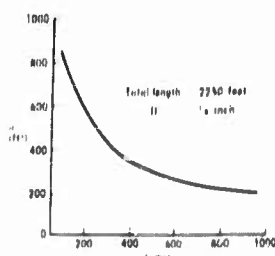
FIGURE 18-3. Rope response versus frequency.

profitable to use tuned rope for a single frequency since so much weight is required to separate the dipoles. However, if a broadband unit consisting of dipoles of various lengths is being considered, it is possible to separate dipoles tuned to the same frequency by using dipoles of other lengths. Only a small amount of the "inert" material need be used. This brings the efficiency to a point where tuned rope may become of use in the low intermediate range from about 100 to 600 megacycles (Reference 11).

18.2.3 Corners

The reflective properties of corners and similar devices turn out to have little application in most ordinary situations calling for confusion reflectors. The reason for this is that the tolerances permitted in the construction of such devices are so small that light, dispensable units cannot be made. If the reflective properties are to be important, the linear dimensions need to be many wavelengths long. Under such circumstances, diffraction effects are small compared to the geometrical effects. It can be shown* that the value of the cross section should be

$$\sigma = 4\pi(A/\lambda)^2 \tag{18-3}$$

The effective area A in this equation can be found by considering the incident wavefront, projecting the corner on the wavefront, projecting the image of the corner in itself on the wavefront, and taking the overlapping area. If the intersection angles of the faces of the corner differ from 90° by an angle of α radians, the cross section will be down 3 decibels if $\alpha = 0.35\lambda/h$, where h is the length of one side of the corner. Because of these rather stringent requirements no really satisfactory dispensable corner has been designed. For this reason the theory will not be presented in any greater detail. Other reflective devices present similar difficulties for use from aircraft.

18.3 Electrical Characteristics

The previous section has summarized the theoretical approach to the determination of some of the electrical characteristics of chaff. All of these theories apply to idealized cases only; the actual situation may be much more complicated. From the standpoint of the chaff package designer, there is little need to look at the electrical characteristics in detail. The user need only proceed under the assumption that each package will provide an adequate echo with any polarization in the frequency region that is appropriate.

*The essence of the proof is the assumption that all energy falling on the area A is reradiated toward the radar as if by an antenna which has a gain appropriate to one with an aperture A. Also see Reference 12.

He should also know what to expect with regard to its mechanical behavior (to be discussed in the next section). For the design of efficient chaff dipoles, the details of the electrical and mechanical behavior are the important factors.

18.3.1 Response versus Frequency

At this time, there is no better theory available than the Van Vleck-Bloch-Hammermesh theory (generally called the Van Vleck theory) for dipoles and the Bloch-Hammermesh-Philips theory for rope; all designs have been based on these theories. There are some indications that the theories underestimate the response of chaff, especially in the nonresonant regions. There have been no definitive experiments, although particular chaff orientations have been studied on back-scattering ranges; some laboratory work with small clouds of dipoles falling in a vertical shaft has also been conducted. Until better information is available, the predictions of the Van Vleck theory are likely to be used for design purposes. It has been customary to perform field tests in which echoes from test units are compared with echoes from units of familiar design as well as with echoes from aircraft. Any effects (such as birdsnesting) that arise from the dispensing process are thus included, and a reasonably significant verification of the proper behavior of the unit is made for several frequencies lying in the nominal range of the material. Long experience with such field tests has led to a rather high degree of confidence in the performance of regular units and in the designer's ability to predict the behavior of more or less conventional units. As newer materials come along, more extensive field tests are called for to establish the behavior of these materials.

The principal problem of the chaff package designer then is placing the proper number of dipoles of the right lengths in each bundle. It has been customary to try to design for approximately uniform response throughout the nominal bandwidth of the bundle. This can be done in a number of ways, such as by using a number of groups of dipoles, each of a different and appropriate length and number, or by cutting the lengths so that there is an almost continuous variation in length, a "diagonal cut." In practice, the first is most commonly used since it is easier to manufacture. The second is used in a few special cases. It is not convenient to design a unit so that the cross section is really uniform with frequency. This is not important since aircraft show a large variation in cross section and since most radars are not very sensitive to minor amplitude variations. It is customary to be content with a variation of about ± 3 decibels in power level from the nominal value. Figure 18-4 shows the behavior of a typical example. This figure has been based upon the Van Vleck theory and is thus subject to whatever limitations exist
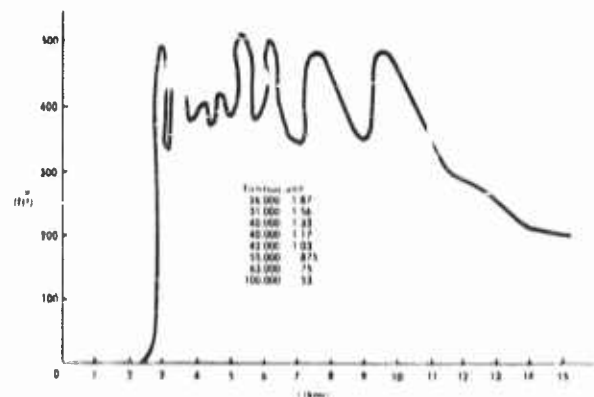
FIGURE 18-4. Response of typical although fictitous unit.

In this theory. As has been pointed out before, there are some reasons to believe that the nonresonant response is underestimated by the theory; so the irregularities may actually not be so great and the response at the higher frequencies may also be underestimated.

As long as the individual dipoles are separated by several wavelengths and are located at random so that the echoes from individual dipoles are noncoherent, one may safely consider the cross section at a given frequency proportional to the number of dipoles. If the density of dipoles becomes such that they are closer together than this, the interaction between near neighbors makes the usual theory inapplicable. In most cases, the effect of interactions tends to decrease the value of σ from that predicted by the theory that neglects interactions. Since the number of dipoles needed to produce a given cross section is roughly proportional to the square of the frequency and the necessary spacing for negligible interaction is inversely proportional to the frequency, it can be seen that the volume of the cloud required to have negligible interaction is inversely proportional to the frequency. Thus the linear dimensions are inversely proportional to the cube root of the frequency. The problem of too high a dipole density, thus, is more likely to arise at low frequencies. For example, at X-band, a volume of about 2400 cubic feet is required for a cross section of 600 square feet; at 300 megacycles, a volume of 72,000 cubic feet. The first is a sphere with a radius of 8 feet; the second needs a radius of 27 feet.

18.3.2 Polarization Effects

The Van Vleck theory is ordinarily used with the assumption that the dipoles are randomly oriented in space. If this were in fact the case, the

polarization of the radar would have no effect on the size of the echo from a chaff cloud. But the aerodynamics of freely falling strips (with or without a V-bend, parallel to the length) does not lead to such a random orientation in many cases. The natural way for a dipole to fall is in a roughly horizontal position. The elevation angle of the radar also affects the size of the observed echo through polarization effects. For an elevation of 90° (sighting directly overhead), no difference would be observed for horizontal dipoles except that there would be a larger than expected echo since there are fewer vertical ones. At low elevations, the response with horizontal polarization will be much greater than for a radar with vertical polarization. Naturally, there is a gradual transition between the two extremes.

Attempts to make the orientation more nearly random have consisted in weighting the end of some or all of the dipoles. There seems to be no known theory for the situation; so what is known is the result of experiment. It is likely that very careful loading of the ends could lead to the equivalent of random orientation, but it may be easier to load some for a definite orientation, leaving all the rest unloaded to fall in a horizontal position. Unfortunately the more vertically oriented dipoles seem to fall faster than the horizontal ones; so there is a tendency for a separation into two clouds falling at different rates.

Two other factors of importance are the turbulence of the air and the distortion of the dipoles. Both turbulence and distortion tend to produce a random polarization. The first is likely to lead to rather uniform polarization soon after dispensing. The second probably accounts for an apparent isotropic behavior of the longer lengths (approximately over 4 inches).

Rope is also polarization sensitive, and its behavior is clearly dependent upon the shape of the streamer as well as the polarization and elevation of the radar. Most ordinary circumstances lead to good echoes for either polarization of the radar.

18.3.3 Noise Spectrum

The echo from a cloud of chaff is the resultant of the return from a large number of individual dipoles. Since the phases of these individual amplitudes are random and constantly changing, the echo shows a marked fluctuation. It is generally assumed that the fluctuations follow a Rayleigh distribution. This means that the fraction of the time that the power lies between P and $P + dP$ is

$$f(P) \, dP = (1/P_0) \exp (- P/P_0) \, dP \qquad (18\text{-}4)$$

The rapidity of variations is not indicated by this law; so it must be con-

sidered separately. It is possible that its spectrum might be used for discrimination. The fluctuation rate is determined principally by two processes. The first of these has to do with the rate at which the chaff cloud spreads laterally as a result of turbulence and irregularities of its fall. The other has to
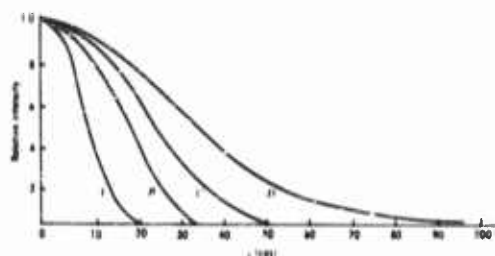
FIGURE 18-5. Frequency spectrums of echo fluctuations.
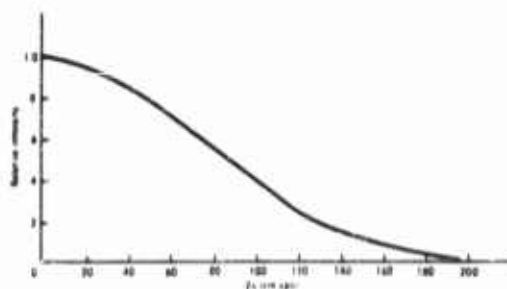
FIGURE 18-6. Normalization of Curve A from Figure 18-5.

do with the rate of rotation of individual dipoles while falling, since this changes the amplitude from each individual dipole. The first changes the phase of each individual amplitude and the second the magnitude of each amplitude. As one would expect, the fluctuations are irregular; so the Fourier analysis will give a very broad spectrum. Figure 18-5 shows such a spectrum based upon measurements made at the Radiation Laboratory during World War II (Reference 13). It will be noticed that the actual spectrum is quite dependent upon the circumstances. Curve A refers to chaff dropped in still air from a blimp. Curves B, C, and D refer to increasingly turbulent conditions, with D for gusty air having velocities varying up to 25 mph. The maximum occurs at a frequency of zero, of course, and falls off in generally the same way. For a given set of conditions, one would expect corresponding signal densities at frequencies proportional to the radio frequency of the radar. It should then be possible to normalize such a curve for all radio

frequencies. Figure 18-6 shows the normalization of curve A from Figure 18-5. If f_0 is the frequency for which $F(f_0)/F(0) = \frac{1}{2}$, it can be seen that

$$f_0 = f_{rf}/375 \quad \text{or} \quad f_0\lambda = 80 \text{ cm cps} \tag{18-5}$$

Here f_{rf} is the radar's radio frequency in megacycles. Under many conditions, f_0 will be somewhat higher, as shown by curves B, C, and D in Figure 18-5. These data are also confirmed by experiments performed by the Radio Research Laboratory during World War II. These measurements were made at 515 megacycles (Reference 14). Since the fluctuations are due to the lateral motion of the dipoles to quite an extent, one should expect the velocity of this lateral motion to be related to the fluctuation frequencies. It can be shown (Reference 13) that, if half the relative velocities lie between $-v$ and $+v$, then

$$v = 0.2\lambda f_0 \tag{18-6}$$

The value of 16 centimeters per second $= \frac{1}{2}$ feet per second is smaller than is usually observed, but it was obtained for very still air and should not be considered typical. The figure at least has the right order of magnitude.

Another way of looking at the same matter is to note that the return at any instant will be correlated to some extent with the return at a previous time. When τ is small, the correlation is very good; when τ is large, the correlation is poor. The correlation will fall off rapidly about where $\tau_0 = 1/f_0$.

The echoes from aircraft show a similar fluctuation, but they differ in several respects. The fluctuations are caused by the changing aircraft aspect with respect to the radar. The change in aspect is due to the aircraft's translational motion and to pitch and yaw. In addition, vibrations of the structure contribute to the fluctuations. The latter effect is much smaller for jet aircraft than for propeller-driven aircraft. Well-known strong components exist in the noise spectrum of such signals at frequencies which are multiples of the propeller frequency. This is called propeller modulation, and is partly due to the effect of the rotation of the propeller blades and partly to the vibrations from the motors. The combined result is that the spectrum of an aircraft echo is nearly a Rayleigh spectrum with f_0 somewhat smaller than for chaff (τ_0 somewhat longer), with certain vibration peaks superimposed. Any of these features may help to distinguish an aircraft from chaff, but this may be quite difficult to accomplish if the echo from the aircraft is immersed in the echo from many bundles of chaff in trail along the radar beam. It is this difficulty that may protect an aircraft from a tracking radar.

18.4 Mechanical Characteristics

The mechanical characteristics of chaff have been quite difficult to determine. This is particularly true of the details of the behavior of individual dipoles; it also applies to some extent to the general motion of chaff clouds. Part of the difficulty is that the behavior is dependent to a considerable extent upon the state of the air in which the chaff is falling. For example, very turbulent air leads to more uniform polarization than quiet air; and measurements of the rate of fall are often confused by vertical air currents (it is not uncommon to find chaff clouds rising rather than falling).

At the present time, one can expect to provide a cross section of 500 square feet at one frequency for about 0.1 pound and 2 to 3 cubic inches. To cover 3 octaves requires about 0.3 pound and 9 cubic inches. These figures do not scale very well, but they give a rough idea of the volumes and weights required. There has been steady progress on this point, and one may expect even higher efficiencies in the future.

The rate at which a unit disperses into a cloud is particularly dependent upon the way the chaff is dispensed. In many cases it spreads almost instantly to a size sufficient to produce a full echo. Under other circumstances it may require up to several seconds for full development of the echo. Rope must unroll to give a full echo, so it requires a noticeable time. Dispensing into a highly turbulent region contributes to the quick development of the echo. After the initial dispensing operation, a free cloud of chaff tends to spread at a rate of about 2 feet per second. If the chaff is in a region in which there are velocity shears, the spread may be quite rapid.

Laboratory studies (References 15, 16, and 17) and field tests show a wide variation for the rate of fall of chaff. Values as low as 120 ± 50 feet per minute to as high as 500 ± 200 feet per minute have been observed at low altitudes. It seems clear that the higher velocities are associated with vertically polarized dipoles and the slower velocities with horizontal dipoles. This makes the clouds tend to separate into a horizontally polarized cloud above a more rapidly falling vertically polarized cloud. Perhaps a figure of 200 to 300 feet per minute is a good value to use for an average figure at low altitudes. Since the viscous drag on an object is given by

$$F_d = \tfrac{1}{2}\rho v^2 C_d A \qquad (18\text{-}7)$$

and the terminal velocity corresponds to

$$F_d = \text{weight} \qquad (18\text{-}8)$$

one can see that the rate of fall should vary inversely as the square root of

the density of the air. Thus at 40,000 feet, where the density is roughly one-fourth that a sea level, the rate of fall should be twice as great. Presumably, the lateral spreading rate will also increase with altitude.

The over-all history of a typical unit of chaff can be summarized as follows:

(a) Upon being dispensed, the package opens almost immediately and a chaff cloud develops. The shape is very dependent upon the exact dispensing conditions. It may have a diameter of about 10 feet and a length of about 50 feet. The echo at S-band and higher frequencies is nearly full size almost immediately; the echo seems to grow around that of the aircraft and separates from it as the chaff is left behind. At L-band and lower frequencies, especially where rope is involved, it may require several seconds for the full echo to develop.

(b) The echo then continues to grow slightly as the chaff cloud spreads out. The cloud may then gradually tend to separate into the two clouds mentioned before, with the horizontally polarized chaff above the vertically polarized material. No very serious change will take place for an interval of about 5 to 10 minutes. The cloud is still smaller than most radar pulse packets.

(c) As time goes on, the cloud may spread in size to several, and possibly many, pulse packets. Thus, its echo is diluted and spread. A vertical separation between the two polarization directions will be significant and indicate that the whole cloud will have fallen through several thousand feet.

(d) After half an hour, the effects of falling and spreading are likely to be so severe and so uncertain that one cannot depend on the chaff at all. However, there have been many cases in which effective clouds of chaff have persisted much longer.

18.5 Materials and Methods of Manufacturing and Dispensing

18.5.1 Standard Materials

The standard material for the production of chaff is aluminum foil with a thickness of about 0.00045 inch. It is cut into strips of various widths such as 0.036 inch, 0.016 inch, and 0.008 inch. Even narrower cuts have been made, but they have not been included in any standard packages yet. Ordinarily, one dipole is given a v-bend to increase its rigidity. This is done by the cutter. The cutting machine is built very much like the cutting portion of a lawn mower. It has a rotating cylinder with knives that cut the foil against a fixed blade. The width of each cut is determined by the feed rate of the foil into the cutting head. The v-bend is produced by having every other blade on the cylinder so dull that it simply bends the foil over without cutting it. Several layers of foil can be cut at the same time. The width of

the web of foil determines the original length of the strips, but the final lengths are determined by forming the correct number of them into a bundle and then cutting them in a guillotine cutter. Difficulties would arise in welding the foil on such cuts of multilayers of foil or in the guillotine cutting process if it were not for the fact that the foil is coated on one side with a thin film of waxes called "slip coating." This also seems to lubricate the knives. Where weighted dipoles are wanted, additional strips of a lacquer loaded with lead powder are printed on the foil. This is called "strip coating." When the foil it cut to length, the lead provides the additional weight to turn the dipoles. Ordinarily the strip coating is put on only one side.

The various bundles of dipoles are kept in paper wrappers after they have been cut to length. They are then assembled in a cardboard sleeve, using the proper proportions of various lengths. The sleeve is designed to be held closed by two tapes which make up a part of the dispensing system. In this system, the tapes are torn off just as the package is dispensed. This permits it to open as it falls and to spill out the dipoles. On occasions, special units are prepared for hand dispensing or for tapeless dispensers. Appropriate sleeves are then used.

Rope is now usually combined with dipoles, although during World War II it was used separately. The standard material now is a roll of plain foil, ¼-inch wide, 0.00045 inch thick, and 750 feet long. The rolls are slit from a wide web and wound on light aluminum ring cores. A light paper tab is attached to the outer end, and the roll is then placed in a light cardboard sleeve. Three or more of these are likely to be used in each unit. Following dispensing, the roll starts unwinding because of the drag of the paper tab. As more and more of the streamer unrolls the drag on the whole streamer supports it and the weight of the core makes the unwinding complete. Even if the foil breaks, it is likely to continue unrolling unless there was severe welding of various layers of the roll.

A diagonal cut of dipoles in the 400-megacycle region can be obtained by making a radial incision, approximately ¼ inch deep, into such a rope roll. Originally, the stock of dipoles so formed was simply left around the rope (of course, enough extra foil was wound to make up for this), but now a separate cut is made and the dipoles are then wrapped separately.

18.5.2 Special Materials

In order to increase the efficiency of chaff, various other materials have been tried. Many of these still show promise for use, and some are being adopted. Examples are given below:

    (a) Glass fibers—coated with silver by chemical methods
    (b) Glass fibers—coated with aluminum by dipping
    (c) Conducting glass fibers

(d) Glass tubes with a conducting core

(e) Fused quartz tubes with a conducting core

(f) Mylar with foil laminated to it

(g) Mylar with metal deposited on it

(h) glass fiber yarn with silver deposited by chemical methods.

(i) Glass fiber yarn spun with a fine wire

The last four materials are used principally in experimental rope. Example (h) has been incorporated into the specifications for the RR-66-AL unit.

18.5.3 Dispensing Systems

The standard dispensing system has been mentioned already. Its action depends upon mounting the packages of chaff on two parallel tapes of reinforced paper tape about 1¼ inches wide, and 4½ inches apart. A single long length of the taped materials containing from 100 to 1000 units can be stored in bins mounted appropriately in the aircraft. If more than the content of one carton is needed, the tapes can be spliced. A stripper mechanism (such as the ALE-6) then pulls the two tapes at the same rate. This is accomplished by a pair of double pulley driven by an electric motor. The motor of the stripper can run at a wide variety of speeds; the dispensing rate can be adjusted to appropriate values. An intervalometer, or other programming device, can be used to provide a programmed dispensing pattern. It can also be arranged so that the dispensing pattern can be chosen manually in flight. As each tape is pulled down by the pulleys of the stripper, they are pulled out of the fastening of the chaff sleeve. The packages then fall down the chute and out of the aircraft. Since the sleeve is no longer fastened, it opens in the slip stream (or even while falling down the chute), releases the chaff. Figure 18-7 is a sketch of the general form of such installations.

Tapeless packaging and dispensing has been provided for certain installations, such as for some of the decoys discussed in Chapter 20. Other special dispensing problems call for special systems. Two notable examples are dispensing from aircraft traveling at speeds in excess of Mach 1 and forward-launched dispensing. No standard installations have so far been determined for dispensing at high velocities, though several experimental systems are
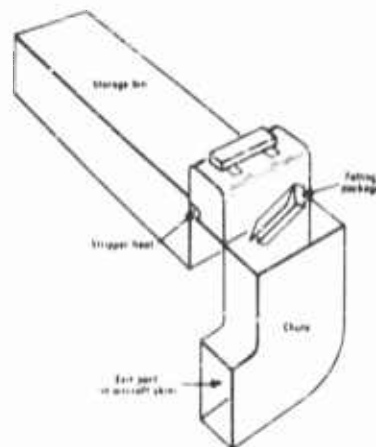
Figure 18-7   Dispenser system

being tried. Forward launching, as discussed in Section 18.6, is accomplished by the use of rockets such as the ALE-9. It consists of a rocket with a payload of about 10 units of X-band chaff packed into a cylinder, about 3 inches in diameter and $2\frac{1}{2}$ feet long, made of plastic. Ten separate charges explode in sequence along the axis of the cylinder, starting at the nose and then dispensing a trail of ten units as the rocket moves ahead of the aircraft. The effect of this launching is discussed later.

18.6 Tactical Considerations

The threat to aircraft that chaff can be expected to combat comes from:
  (a) Radar-controlled antiaircraft guns
  (b) Radar-controlled ground-to-air missiles
  (c) Radar active-seeking missiles
  (d) Fighter aircraft making use of:
      (1) Ground control based on surveillance radar information
      (2) Airborne intercept radar equipment

The equipment involved in (a), (b), (c) and $(d\text{-}2)$ fall into the class of tracking radars; the equipment for $(d\text{-}1)$ is of the scan or surveillance type. Perhaps one should also point out that even local defense using tracking equipment require the background information from surveillance-type radars for full effectiveness since proper target assignment can seldom be made locally and since acquisition of a target is difficult without "setting on" information from scanning type equipment.

It will first be considered how chaff may be used against tracking radars. The classic situation for the use of chaff considers either a conical scan or monopulse radar of high resolution (about 20 yards in range and $\frac{1}{2}$ to 2 mils in both azimuth and elevation). Such high resolving power does not properly indicate the size of the pulse packet. It is likely to be several degrees wide by 100 yards or so in range. A chaff unit dispensed within such a cell will produce an echo merging with that of any other target within the same pulse packet. A trail of chaff units spaced not much more than a pulse apart produces a set of echos in trail from which it is difficult for a radar to separate any normal target. In the ideal situation an aircraft following in a chaff trail already laid would not be resolved, and good tracking information for antiaircraft guns or for missile guidance could not be obtained. In many respects, this is equivalent to other types of noise introduced into the radar signal. The situation is also not really changed when the radar is in a moving vehicle such as in the case of an AI equipped fighter or a homing missile.

Perhaps the principal weakness of this situation it that it is so difficult to realize. Some preceding dispensing vehicle is required. Once the trail is provided it may become very difficult to fly in it or close enough to it to obtain protection. This may be the result of navigational error or simply because the assigned mission may not permit it. The trail alone seems to offer only

limited protection to the dispensing aircraft. However, there seem to be numerous cases in which a sufficient degree of confusion was introduced to seriously degrade the action of the radar and it is probable that the trail of chaff will produce a bias in the indicated position of the aircraft unless an effort is made to overcome the effect. If the operator is using manual or aided tracking, he need only track the leading edge. In automatic tracking devices, simple circuit modifications can accomplish the same end. This is perhaps the simplest form of antijamming.

The problem of protecting a single aircraft from a tracking radar thus poses a problem and some sort of forward launching seems called for. Such a system launches a rocket in the direction of the flight path. The effect to
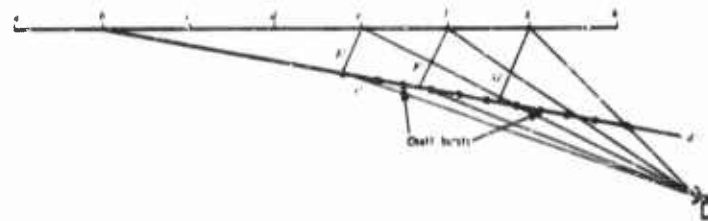
FIGURE 18-8.   Geometric relation of forward-launched chaff to aircraft and radar.

be expected is indicated by Figure 18-8. The line $abcdefg$ is the path of the aircraft with b', c', d' the path of the rocket at corresponding times. The dots between c' and d' indicate the separate bursts of chaff. The radar does not note the chaff until the aircraft is at e. It then gets the chaff signal along with the aircraft; and under the conditions that (1) the chaff is in the range gate, (2) the chaff echo is larger than the aircraft echo and, (3) the distance between the aircraft path at point c' is less than one-half the beamwidth of the antenna and at point d' is greater than the beamwidth of the antenna, the radar concludes that the aircraft is at E'. The rest of the way along the path the radar tracks the position of the aircraft incorrectly. When the aircraft is actually at f, the radar track is at F' and when it gets to g, the radar tracks it at G'. By the time the aircraft reaches h, the radar will have lost it completely; i.e., the chaff may have produced a break-lock.

The effect of chaff on surveillance type radars is next in order. It is difficult to provide enough chaff to screen an aircraft completely. To do this, there must be a unit of chaff in each resolution cell of the radar as for a tracking radar, and, in addition, the chaff should be spread in width as well as in length. This means many more units are required for surveillance radars than for tracking radars. One factor working in the favor of chaff

however, is that surveillance radars usually have much larger pulse packets. In addition they seldom have good height discrimination (unless height finders are added); it is not necessary therefore that the aircraft be at the same altitude as the chaff.

Rather than hope for complete screening, it seems more likely that protection can be obtained by simply confusing the air situation as seen on the PPI scope of the radar. For maximum effect there should be a sufficient number of chaff echoes to saturate the discrimination power of the radar system and operators. Each echo must be examined in some way or other from sweep to sweep to identify proper targets. Certainly the presence of many chaff targets can only make the normal GCI procedure more difficult to carry out, thus degrading the system. Some sort of randomness in position is especially desirable, but hard to accomplish. If one could provide unmamed vehicles to sow in advance, as proposed by Hult (Reference 18), or if a previous wave of aircraft could sow confusing patterns of chaff for a later wave of aircraft, a very high degree of protection could be obtained. Perhaps this is one of the useful functions of decoy systems yet to be discussed.

Another type of protection obtained is concealment of the size of an attacking force. A few aircraft may be made to simulate a much larger strike; and, conversely, if a pattern of heavy chaff sowing has been followed, a reduction in sowing rate may give a picture of a smaller force than usual. If the enemy overestimates the strength of a striking force, he may commit a much larger portion of his defensive strength than he should. The enemy may thus be left too weak to make a proper response to the main force coming later.

Before proceeding to the track-while-scan systems it is well to point out that a single aircraft dropping chaff may only increase its chance of detection at maximum range and then mark its trail for the enemy. But it may make the enemy apprehensive of other arriving aircraft and he may hesitate before assigning the proper force against the one target. This itself may make the chaff worth while.

Track-while-scan systems are beginning to appear as part of the radar technique. In these systems, the data presented for over-all surveillance observations are given in such detail and with such resolution that useful tracking information can be obtained. It is possible to use lock-on techniques, and a system of this sort can track simultaneously quite a number of separate targets. Usually the precision of such tracking is not as great as for a standard tracking system, but it may still be quite enough for missile guidance. Of course, the same sort of screening that will work with a tracking radar will work with a track-while-scan system. The decreased resolving power may even call for smaller amounts of chaff. But even if a proper trail cannot be provided, all extra targets will help degrade the system since they must be

examined at least once before they are discarded. If enough chaff can be provided, serious degradation may occur.

There is no attempt to treat all situations or geometries in this section. Single aircraft, multiple aircraft, succeeding waves of aircraft, varying courses, changing response by the enemy, and many other factors provide for many interesting opportunities for the use of chaff. The reader may well wish to think out what will happen under some other set of circumstances than those indicated here. A simple example might be the case of a beam or near beam approach by a fighter with AI equipment. Many other examples can be thought up easily. It is hoped that several such exercises will give the reader a chance to see why the term "weapon of opportunity" is used for chaff.

One final point for this section is to consider the compatability of chaff and jamming. It should seem clear that, in general, the use of one will not hurt the use of the other. One can think up examples in which this is not the case, but, in the actual environment of conflict these special cases are not likely to ocurr nor are the most effective anti-jamming measures likely to be taken. In any case it is clear that the broadband possibilities of chaff help to discourage wide-range frequency shifting, thus backing up any jamming program. Conversely, the threat of jamming tends to discourage a number of antichaff systems. In each individual case the combination is very likely to be an improvement over either one.

18.7 Antijamming

The problem of antijamming (A-J) of chaff can be thought of either as discrimination between chaff and target or elimination of the chaff "noise." The two are essentially equivalent. It is easiest to describe procedure in terms of discrimination, but this is only a matter of convenience. The basis of discrimination may be any feature of the chaff echo which differs from that of the target. A number of such features are:

  (a) Lack of motion
  (b) Different fluctuation characteristics
  (c) Different polarization characteristics
  (d) Different echo size
  (e) Spatial position

How such a lack of equivalence to an aircraft can be used depends in detail upon the radar system. Where the system normally uses an operator all of the fine details of differences observable by a person may be put to use. An individual can do a great deal of integration and sorting and it is almost a truism that "a well-trained operator is the very best A-J device of all against chaff." But it is possible to saturate the sorting ability of even the

best-trained operator. The differences may also be sufficiently obscure so
that the delays in discrimination may amount to a serious degradation of the
system. Of course, in the case of extreme difference echo size an operator
will have little difficulty in picking out the chaff. Items (b), (c), and (d)
should in principle furnish useful discrimination, but it has not been generally
practicable to depend on them except in the case just mentioned above. One
other case in which the echo size may play a role occurs when the dipole
orientation is so poor that a very weak return is obtained. The lack of motion
of chaff or its special position have seemed to present the most useful weak-
ness to exploit for A-J purposes.

A number of schemes for moving target indication (MTI) are in use. The
common ones depend upon pulse-to-pulse cancellation of fixed targets. In
order to have some difference for moving targets so that the cancellation is
incomplete, it is customary to use the doppler shift in frequency resulting
from some radial component of velocity. If the returned signal is mixed with
one exactly like the transmitted signal, a beat will result. This fluctuation of
signal makes the substraction of two sequential sets of echo give a non-zero
result for moving targets and a zero result for stationary targets. The com-
parison signal may be obtained by using the continuously running local oscil-
lator to control the transmitter frequency, or it may be obtained by using a
reflected signal from a stationary target. The first type is referred to as a
coherent MTI and the other is called a noncoherent MTI. The latter can
only work if there is a stationary target to provide the comparison signal.
Chaff itself might provide this target. Some radars can be made to change
from coherent to noncoherent operation and back at will.

Such MTI systems are not perfect, though they may provide significant
and useful discrimination. Any tangentially moving target is essentially a
fixed target and will not be seen. There are also certain radial velocities for
which no target will be seen. These "blind velocities" are those for which the
beat frequency is an integral multiple of the pulse repetition frequency. Then
the beat produces no change in pulse shape and there will be full cancellation.
These velocities are those for which the frequency shift $\triangle v = np$, where p
is the prf and n is an integer. If v_0 is the radio frequency of the radar, v_r is
the radial velocity of the target, and c is the velocity of a radar signal then

$$\triangle v = v_0 v_r / c \qquad (18\text{-}9)$$

Thus the blind velocities are

$$v_r = \frac{np}{v_0} c \qquad (18\text{-}10)$$

To get a feel for these speeds note that for S-band ($v_0 = 3 \times 10^9$ cps) and

a prf of 500 per second the first blind velocity is about 100 knots. It is desirable to keep n as small as possible, that is, to make the blind speeds as high as possible. This may be done by increasing the prf or by decreasing the radio frequency. If the first is done, it shortens the unambiguous range, and if the second is done, it decreases the resolving power with the same size antenna. Thus the need to counter chaff calls for measures which degrade the quality of the radar. In addition, the problem of winds aloft may make it impossible to eliminate the chaff targets. These wind velocities may be quite high. Besides this, they may be different at different altitudes, which prevents any compensation scheme from being universally effective. It should be clear that the nature of chaff does not make it useful against a continuous-wave doppler system.

The situations discussed are nearly all dependent upon the effects of the atmosphere. In the case of operations outside the atmosphere, these effects no longer hold or are changed markedly. Even the lightest chaff tends to continue in a free orbit at full orbital velocities. The very slight drag produced by upper remnants of the atmosphere can show itself only as a very small effect that can be observed only as it accumulates over a long period of time.

REFERENCES

1. L. J. Chu, *Analysis of Window and Related Matters*, Report 4, Radio Research Laboratory, Harvard University, October 1942.

2. J. H. Van Vleck, F. Bloch, and M. Hammermesh, "Theory of Radar Reflection from Wires or Thin Metallic Strips," *J. Appl. Phys.*, Vol. 18, p. 274, 1947.

3. *Interim Engineering Report*, Standard Rolling Mills, Inc., Report No. 3-6, March 1954.

4. *Theory of Electromagnetic Waves* (A Symposium), Interscience, New York, 1951.

5. C. T. Tai, Stanford Research Institute, Technical Report No. 18, 1951.

6. J. Sevick and J. E. Storer, Cruft Laboratory, Technical Report No. 149, 1952.

7. J. Sevick, Cruft Laboratory, Technical Report No. 150, 1952.

8. R. C. Kouyoumjian, The Ohio State University Antenna Laboratory, Technical Report No. 444-13, 1953.

9. J. M. Minkowski, and Edwin Cassedy, "Cross Section of Collinear Arrays at Normal Incidence," *J. Appl. Phys.*, Vol. 27, p. 313, 1956.

10. F. Bloch, M. Hammermesh, and M. Phillips, "Radar Reflection from Long Conductors," *J. Appl. Phys.*, Vol. 17, p. 1015, 1946.

11. F. Cassedy, J. M. Cotton, and J. M. Minkowski, *A Summary of Research on Tuned Rope*, The Johns Hopkins University Radiation Laboratory, Technical Report No. AF-46, February 1958.

12. J. L. Lawson and G. E. Uhlenbeck, *Radar System Engineering*, Radiation Laboratory Series, Vol. 1, McGraw-Hill, New York, p. 167.

13. J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals*, Radiation Laboratory Series, Vol. 24, McGraw-Hill,, New York, p. 133.

14. G. P. Kuiper, *A Study of Chaff Echoes at 515 Mc*, Radio Research Laboratory, Harvard University, Report No. 411-73.

15. G. G. Anthony et al., *Characteristics and Uses of Chaff*, Halier, Raymond and Brown, Inc., Report No. 29F.

16. *Interim Engineering Report 4-4*, Revere Copper and Brass Co., Foil Division, p. 9, Table PT 60-123.

17. *Interim Engineering Report 3-8*, Revere Copper and Brass Co., Foil Division, p. 35.

18. J. L. Hult, *Electronic Countermeasures Against Air Defense Systems*, The RAND Corporation, Report No. RM-1090, 1953.

19

Target Masking and Modification

J. H. VOGELMAN,   R. K. THOMAS

19.1 General Considerations

19.1.1 Nature of the Problem

This chapter deals with appropriate techniques for artificially adjusting the magnitude of the parameter σ, the echoing area or radar cross section, which occurs in the radar equation.

Here σ is defined in the conventional manner to be the area of an isotropic scatterer which would produce the same echo as the actual target in the direction of the radiation source. This is the monostatic case, as distinguished from the bistatic one for which the radar return of interest is in a direction other than that of the source.

Due to the complex nature of most target structures of interest, σ will vary over a considerable range as a function of viewing angle and polarization. For the purposes of this chapter, unless otherwise stated, these quantities will be considered as constant in the evaluation of changes in σ due to the techniques under consideration. Since the maximum range of a given radar is proportional to the fourth root of σ for a nonextended target, it is evident that large changes are required in this parameter in

order to produce a significant effect. A change may be made in either direction, depending upon the nature of the desired result. In those situations where the intent is to elude radar detection by becoming "invisible", the change must lead of course to a reduced value of the echoing area. There are circumstances, however, when, in order to alter the natural radar presentation, it is desirable to increase the echoing area so that targets will appear where none would normally show. For example, for non-normal angles of incidence, the echoing area of bodies of water is very small compared to a land return. Bodies of water appear as blanks on the radar scope, and often serve as navigational aids. If the echoing area is properly enhanced in these regions, the radar return level relative to that of land will be increased and the navigation aid would be removed.

The means by which the echoing area may be reduced are purely passive in character, whereas the methods for increasing the echo may be either passive or active. In the active case, the technique centers around the electronic generation of signals shaped and timed to correspond to the signal characteristics of the illuminating radar in order to preserve a natural appearance on the scope. The required increase in σ is obtained by choosing an appropriate power level. In the passive case, the returned signal is that of the originating radar and the entire emphasis is on concentrating the reflection to as high an extent as is possible. Insofar as target reduction is concerned, the method consists of absorbing or preventing a coherent return of the incident radiation.

19.1.2 Distinction between Ground and Airborne Target Masking

Since the background of an airborne target is quite different from that of a ground target, a difference in technique may be expected in the detection and prevention of detection for the two cases. The airborne target is comparatively isolated in free space, and there are no competing returns, in the absence of extraneous decoys, to hamper the identification of the important target. The use of such schemes is covered in Chapter 20; for the purposes of the present chapter, all airborne targets will be considered to be genuine offensive weapons for which a reduction in radar cross section would be a desirable advantage. The design objective in this connection would be to minimize this quantity to the greatest practicable extent, the limitations being due to the size and weight of the absorbing material required to do the job. With the present state-of-the-art, adequate broadband protection would impose a severe penalty on the flight characteristics of a large aircraft if the attempt were made to cover it with absorbing ma-

terial. The use of structural material having absorbent properties is an-
other matter, and will be treated in Section 19.3.3 of this chapter.

It should be noted that an ultimate limitation on the radar cross section
of an airborne target is set by the electronic transmitting and receiving
equipment on the aircraft. Radar and communications antennas, which of
course cannot be covered with absorbing material, create a residual cross
section of about ten square meters. This would be the irreducible minimum
value of σ, assuming perfect absorption everywhere else.

Airborne target reduction must be effective against bistatic radars and the
absorbing material, consequently, must preserve its electrical characteristics
over a wider angular region than would be required for the protection of
ground targets.

In the absence of absorbing materials, the scattering from an airframe
can be divided into contributions from three types of surfaces: 1) doubly
curved, 2) singly curved (cylindrical), and 3) flat surfaces. When the
radii of curvature and surface dimensions are large relative to the incident
wavelength, the scattering becomes more sharply focused. In the limit of
geometrical optics, purely specular reflection is observed.

An interesting technique for altering the cross section of an airborne target
consists of modifying the surface configuration according to the principles
of geometrical optics in such a manner that large reflections are diverted
to unimportant regions of space at no increase in magnitude. By this means,
as a typical example, it has been possible to obtain a 10 to 1 reduction
in broadside cross section in the horizontal plane at the expense of three per-
cent increase in frontal area. The aircraft surfaces are altered so that the
incident radiation is reflected at angles of about 20 degrees above and below
the horizon.

The data taken in this connection were obtained by optical techniques on
scale models and have been found in good agreement with calculated values
as well as those measured by true scaling of model and wavelength. The
10-decibel figure mentioned above corresponds to the practical modification
of a conventional aircraft. Being based on the limit of geometrical optics, this
approach is of course only valid for wavelengths very small compared to
airplane dimensions.

The surfaces of airframes are prescribed by aerodynamics and hence the
change in contour, i.e., conversions to flat surfaces, required by the method
mentioned above find little application. However, in the absence of re-
strictions on allowable orientations and shapes, the control of scattering
by reflecting surfaces may be competitive with control by absorbers.

Ground targets are situated in an environment which generally produces a variety of reflections, one important exception occurring when the "ground" is the surface of the sea. Since the desired objective is to render the target as inconspicuous as possible, the means of accomplishing this will usually be more complex than the mere reduction of the radar cross section of each component of the target. To blend in with the surrounding terrain, various portions of the target may require no treatment or enhancement, either passive or active. The situation is quite different from that of an airborne target, which stands out against an empty background.

19.1.3 Limitations on Size of Target

As stated previously, all other factors remaining equal, the range of a given radar is proportional to the fourth root of σ, the radar cross section. There may be cases, however, where the natural value of σ is so enormous that, even when substantially reduced by the various available techniques, the resulting radar range is more than adequate. As an example of this, consider a large ship which would have a radar cross section on the order of one million square meters, with the lack of reflection from the surrounding water providing a high degree of contrast. Suppose that it were possible to reduce the cross section by 20 decibels, which would be a difficult task to perform, even over a relatively narrow band. Using the reduced value of $\sigma = 10^4$ square meters in the radar equation, along with nominal values of 60 kilowatts peak power, 30 decibels antenna gain, and 84 decibels below a milliwatt (-dbm) receiver sensitivity at X-band yields a range of approximately 90 kilometers. An airborne radar having these characteristics would thus retain a considerable capability against a ship protected to the extent indicated above, and reducing the ship's cross section would be of little value.

The foregoing example indicates a limitation on the usefulness of σ-reduction in the case where large targets are involved. The case chosen to illustrate this is extreme, and the same considerations would not necessarily apply to land based targets, for reasons previously mentioned.

19.2 Methods of Target Masking and Modification

19.2.1 Absorbers

Insofar as target reduction is concerned, the ideal countermeasure is one which renders a target completely invisible to an enemy radar. This

may be accomplished in theory by employing a suitable absorbing material as a cover which simulates the impedance of free space. In practice, this is a difficult technique to apply, especially, if the wavelengths under consideration are to range over a wide band, as is usually the case. The progress which has been made to date in developing absorbing materials is reported in subsequent paragraphs of this chapter; it will be seen that significant broadband reductions in σ may be achieved with small weight and large volume, or large weight and small volume, but the combination of small weight and small volume has yet to be realized. This is not true for narrow-band applications, where a satisfactory combination of electrical and mechanical properties may be obtained quite readily. This is an important point, since not all operational requirements call for a wide frequency band. Certain classes of targets will be subjected only to high-resolution radar operating in the microwave bands and may be protected at little sacrifice in weight or volume. There are, on the other hand, targets for which mechanical restrictions are relatively unimportant and in this case good broadband protection can be provided.

19.2.2 Reflectors

Reflectors are passive devices intended to augment the radar return from a given area and, when used in combination with absorbers, serve to alter the normal radar presentation thus providing an electronic camouflage. The function of a reflector is, of course, exactly the opposite of an absorber, and it should operate in this fashion over the same bandwidth as the absorber and for the same viewing angles.

For all practical forms of reflectors, there is a natural dependence of σ upon the reflector dimensions expressed in wavelengths, so that a given physical shape will have bandwidth properties beyond the control of the designer. The viewing angle over which effective reflection may occur is, on the other hand, subject to a large measure of control, depending upon the extent to which the mechanical consequences may be tolerated. A flat metal plate is the simplest configuration; it yields a high reflection over a narrow angle. There are various forms of corner reflectors, which are three-dimensional objects, capable of operating over wider angles. Still wider angle performance may be obtained through the use of a properly arranged lens such as the Luneberg type, which represents another degree of constructional complexity. Design relationships for these reflectors are given in section 19.4, which will assist in the selection of the best configuration for a particular application.

19.2.3 Structure and Terrain Centouring

This topic deals with the electronic camouflage of land targets and surrounding areas by means of the absorption-reflection techniques discussed above. The quality of the results depends on the material and effort which can be allocated to the purpose, since weight and volume restrictions are not of prime importance here. Broadband absorbing material is commercially available, or can be produced readily for a variety of applications, including those requiring a substantial degree of weather resistance. As previously indicated, the reduction in cross section need not necessarily be great for every part of the target; various regions of it will require differing degrees of protection, which is of importance from an economic point of view.

Insofar as reflecting elements are concerned, corner reflectors would be the most economic type if large quantities were required. This will generally be the case since important target areas or their simulated counterparts are extensive in area. Not only may real targets be altered by these means but false targets may also be created to confuse the radar operator who depends on the presence of real targets to fulfill the navigational requirements of his mission. Some applications of this technique will be discussed in section 19.5 of this chapter.

19.2.4 Electronic Target Simulation

Electronic target simulation refers to the generation of false targets by active means. There is a distinction between this technique and distributed area jamming. The latter method employs large numbers of jamming transmitters located throughout a large ground area and produces many sources of masking radiation which are not intended to create the appearance of genuine targets. The radar operator will know that he is being jammed and will be faced with a problem different from the one posed by target simulation.

In the case of target simulation, the objective is to produce a radar return which will not contain information that might reveal its spurious nature, being merely a stronger echo of that which would normally occur. A typical application of this technique would be to raise the apparent strength of naval units by giving small vessels carrying the appropriate equipment the electronic dimensions of destroyers, or battleships. A similar concept may be employed in the case of airborne vehicles; this situation is treated at some length in Chapter 20, "Decoys".

In accordance with the foregoing, a target simulator consists of equipment which receives the radar signal and retransmits a faithful duplication

of it with a minimum of time delay. The apparent size of the target would depend upon the power employed; the power need not be excessive to create significant effects. Omnidirectional antennas are generally required, and the system should be relatively broadband to cope with a variety of radars. Difficulties will naturally arise if the number of these becomes too great since one equipment cannot provide a satisfactory reply to an overwhelming quantity of interrogations. The general characteristics of this type of equipment are discussed in section 19.6.

19.3 Properties and Applications of Absorbent Materials

19.3.1 Broadband Compositions

Any absorbent material may be analyzed as a section of a transmission line having a characteristic impedance given by the usual relationship

$$Z_0 = \sqrt{\frac{\mu}{\epsilon}}$$

where μ and ϵ are complex quantities expressed as $\mu = \mu' - i\mu''$ and $\epsilon = \epsilon' - i\epsilon''$. The loss tangents are μ''/μ' and ϵ''/ϵ', and these should be high for rapid attenuation in the "line". The lower these values are, the greater the line length, or material thickness in this case, must be to produce a given attenuation.

The equivalent circuit for a broadband absorber can be considered as a transmission line of constant characteristic impedance matching free space, composed of lossy elements, or as a transmission line of tapering characteristic impedance such that its input impedance is that of free space at the surface, transforming down to a perfect conductor or short-circuit.

Several versions of broadband absorbers are commercially available, consisting generally of a relatively light-weight foam or hair mat, impregnated with lossy material such as conductive rubber. The required thickness depends upon the maximum wavelength for which the absorber is to function; the theoretical minimum would be about $0.1\lambda_{max}$ for a homogeneous composition depending upon the exact absorption specifications. Values of $0.15\lambda_{max}$ have been obtained in practice.

Table 19-I lists the properties of a few typical absorbers of this class, all
having a power absorption of 98 percent. This value is based on normal

TABLE 19-I. TYPES OF ABSORBERS

Type	Lowest frequency (kmc)	Thickness (inches)	Weight (lb/ft²)
Hair mat	2.5	2	0.33
Flexible foam	2.4	1	0.50
Rigid plastic foam	2.3	2	0.62
Hair mat (pyramids)	0.15	26	3
Flexible foam (pyramids)	0.10	36	2.75
Rigid plastic foam (pyramids)	0.04	48	5

incidence of the wavefront, and it should be emphasized that for non-normal
incidence, the characteristics may differ. A good quality absorber should,
however, operate satisfactorily for all polarizations with incidence angles
up to about 30 degrees. Since reflections at such angles will not return to
the source, any decrease in performance in this respect is of no consequence
insofar as monostatic radar is concerned.

These absorbers retain their properties throughout the microwave region,
and, hence, are extremely broadband. The last three items in Table 19-I have
a surface formed in the shape of pyramids. This provides a more gradual
transition from free space to the absorbing material, thus producing less
reflection. It is seen that these absorbers are relatively low in weight, at the
expense of being very bulky, and are not suited for aircraft application. They
are applicable, however, to fixed structures and are widely used for reducing
reflections in test sites. Compared to the hair mat, the plastic material has
superior weather-resistant properties.

It is possible to reduce the bulk to a considerable extent at the expense
of increased weight through the use of ferrite materials. The previously
quoted minimum thickness of $0.1\lambda_{min}$ was based on the use of nonmagnetic
substances; it may be reduced by a factor of $|\mu|/\mu_0$ through the use of a
suitable magnetic material in the case of nonresonant configurations. Ferrite
compositions appear to be the most promising in this respect. They involve
high-loss magnetic materials embedded in a dielectric matrix.

Relatively thin resonant configurations based upon the use of ferrites are available. They exhibit bandwidths on the order of 10 to 1. This appears

TABLE 19-II. FERRITES CAPABLE OF 95 PERCENT ABSORPTION
AT NORMAL INCIDENCE

Type	Lowest frequency (kmc)	Bandwidth (ratio)	Thickness (inches)	Weight (ib/ft²)
Ni-Mn-Zn	0.10	13.2	0.217	5.20
Ni-Zn	0.05	11.2	0.347	8.32

to be exceptional inasmuch as resonant-type absorbers generally have fractional bandwidths (see next section). Resonant ferrite structures yield such bandwidths because of the manner in which ϵ and μ vary with frequency. Effectively, they cause a given physical thickness to maintain an electrical thickness of a quarter wave over an extended region of wavelengths. Table 19-II lists two such ferrite materials, capable of 95 percent absorption at normal incidence.

Although these compositions require little volume, their weight is excessive for airborne applications. Heating also constitutes a problem since the Curie temperature at which the magnetic characteristics disappear is relatively low. These materials, at present, seem best suited to fixed-target or small non-airborne-target covering. Their weather resistance is superior to that of the hair mat or foam type materials.

19.3.2 Narrow-Band Compositions

Narrow-band absorbers include a wide class of special impedance-matching structures composed of layers of different materials, the necessary loss being provided by resistive elements or sheets. They can be made quite thin compared to the broadband types which are essentially homogeneous absorbers, the equivalent circuit for which is a transmission line of constant, or possibly tapering, characteristic impedance.

In the narrow-band case, the equivalent circ consists of a number of transmission lines connected in series. Each lay of the material may be considered as one such line, and its constants are adjusted to provide a certain impedance transformation from the preceding layer. By this means, one may start from the normally reflecting target surface and, through a series of transformations, arrive at a surface impedance which matches that of free space, the necessary attenuation having been included in the

design of some or all of the material layers. The number of parameters available for adjustment is thus larger than in the case of the homogeneous broadband absorber; this additional complication manifests itself electrically as a resonance phenomenon having typically narrow-band characteristics.

During World War II, two principal types of resonant absorbers were developed by German scientists and used with success in snorkel camouflage. The first of these, the Jaumann absorber, consisted of seven layers of thin semiconducting sheets, separated by dielectric layers of equal thickness, about 0.35 inch. The lossy sheets were made of lampblack-impregnated paper having a thickness of 0.004 inch, and the conductivity of the various sheets was graduated exponentially from layer to layer. The second absorber, known as "Wesch-Mat", consisted of two layers with a total thickness of about 0.3 inch, which was considerably less than the Jaumann absorber purporting a similar wavelength coverage. The first layer consisted of a 0.04 inch thick rubber sheet. This was covered by a mat of a rubber-like substance, molded in the form of a waffle and loaded with carbonyl iron powder. Good absorbing properties were claimed at 3.3 and 10 kilomegacycles, with a resonant reflection occurring at 6 kilomegacycles.

Subsequent work in Germany has led to the development of a technique which consists of spraying a metal surface with eight coats of a paint-like material, each having different electromagnetic properties. This process provides excellent environmental and mechanical characteristics, which make it suitable for application to aircraft. The total thickness is about 0.08 inch, with a weight of about 0.76 pound per square foot. The material absorbs 93 percent or more energy over a 20 percent bandwidth in the 10 kilomegacycle region.

Research as been initiated with the objective of increased bandwidth by investigating such items as electric and magnetic dipole absorbers and absorption of energy in anisotropic media. Emphasis in the first category has been placed upon absorbers incorporating resonant elements or loops either arranged within the metal backing sheet itself or by specially dimensioned resonant circuits. Preliminary results indicate an absorption of better than 95 percent in the 2.5 to 6 kilomegacycle band for a material thickness of approximately 0.8 inch. Insofar as anisotropic compositions are concerned, suitable ferromagnetic materials have yet to be developed, but it has been predicted that an absorber thickness of 0.033 inch over a metal surface or 0.010 inch over a masonry surface might be produced, providing 98 percent absorption in the 5 to 38 kilomegacycle frequency band.

As previously noted, resonance absorbers involving ferrite materials in general have superior bandwidth properties because of the manner in which μ and ϵ can vary with frequency.

Absorbent coatings for airborne use must be thin, lightweight, and durable. Presently the only available materials which meet such specifications are relatively narrow-band and function in the microwave frequency regions. Insofar as their application is concerned, it should be noted that certain aircraft portions contribute more than others to the radar cross section; for example, the head-on return is principally due to wing edges and engine nacelles. Thus, selective covering must be employed and installed in accordance with experimental measurements for each given type of aircraft.

19.3.3 Structural Absorbers

In contrast to the techniques previously discussed, which deal with the covering of reflecting structures, there remains the use of structural material with inherent absorbing properties. Such materials would be suitable for both airborne and ground configurations and have been the subject of continuing research for some time. In one version, the absorber has a honeycomb construction, with loss provided by incorporating material of logarithmically graded resistance along the sides of the honeycomb cells. The dielectric material consists of phenolic-fiberglas, both for the facing sheets and honeycomb, and the resistive material is a combination of carbon black and silver powder. Absorption of greater than 95 percent over a 2.5 to 13 kilomegacycle range has been achieved with a material thickness of one inch and weighing 0.7 pound per square foot, which is satisfactory for some subsonic vehicle designs. A material with mechanical characteristics suitable to vehicles operating at supersonic speeds remains to be developed; the associated high temperatures and erosion processes present serious problems.

19.4 The Design and Use of Reflectors

19.4.1 Corner Reflectors

Figure 19-1 shows three types of corner reflectors. All consist of mutually perpendicular plane surfaces, meeting in an apex denoted by O in the sketch, which represents the appearance of the reflector as viewed along the axis of symmetry. Triangular corner reflectors are those having sides $AA'BO$, circular corner reflectors have sides $AA''BO$, and square corner reflectors

have sides $AA'''BO$. The projected areas of each would be as shown; the reflection obtained in each case depends directly upon the projected area. Formulas are given in Table 19-III for the various maximum radar cross sections and that of a flat plate. The maximum value of σ occurs along the axis of symmetry, which in the case of a flat plate is the perpendicular to its surface.

TABLE 19-III. FORMULAS FOR RADAR
CROSS SECTION AND A FLAT PLATE

Reflector Type	σ_{max}	
Flat Plate	$1.000 \left(\dfrac{x}{\lambda} \right)^2$	4π
Square Corner	$1.000 \left(\dfrac{x}{\lambda} \right)^2$	4π
Circular Corner	$0.718 \left(\dfrac{x}{\lambda} \right)^2$	4π
Triangular Corner	$0.445 \left(\dfrac{x}{\lambda} \right)^2$	4π

$\lambda =$ free-space wavelength

$x =$ projected area according to the shape of the reflector as shown in Figure 19-1.
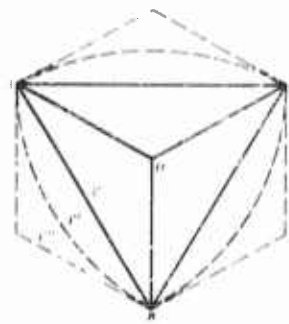
Projected area of:	Value is:
Triangular corner is within 1';	0.433 ($l/h)^2$
Circular corner is within .4";	0.680 ($l/h)^2$
Square corner is within 1''';	0.966 ($l/h)^2$

FIGURE 19-1 Corner Reflectors
(projected view)

The values are arranged in a descending order of magnitude; it is seen that the flat plate provides the greatest return for a given projected area. This is true, however, for only a very limited angular region; the flat plate is therefore unsuitable if any appreciable range of viewing angles is involved. For example, at a viewing angle of less than two degrees from the perpendicular, a flat plate 10 wavelengths on a side has a radar cross section that is one tenth of the maximum value. In comparison, the cross section of a triangular corner reflector is only reduced to one half of the maximum at an angle of 20 degrees. Operational requirements generally call for wide-angle coverage, and the three types of corners differ little in this respect. The triangular version, although having the lowest value of σ_{max}, is generally chosen because it can be easily manufactured and handled. It should be noted that the radar cross section in all cases is a function of the wavelength. This is the only additional electrical design factor which must be considered, other than that of insuring that highly conducting materials are used in order to maximize reflections.

19.4.2 Luneberg Lens

In the event that wider angle coverage is desired than may be obtained from a corner reflector, a suitable means of achieving this result is the use of a Luneberg lens, applications of which are treated in Chapter 10. In its simplest form, this lens consists of a sphere constructed of low-loss material having a value of ϵ which varies with the distance from the center of the sphere in accordance with the relationship

$$\epsilon(r) = 2 - (r/R)^2$$

where R is the radius of the sphere. This creates a lens action which brings a plane wave incident upon the sphere to a focus on the surface at a point diametrically opposite from the point of tangency of the wavefront. Thus, if a small reflecting disk is located at any point on the lens surface, plane reflection from the projected area of the sphere results along the direction of the line joining this disk and the center of the lens. If a larger portion of the sphere is covered with a reflecting material, the same magnitude of reflection results over all viewing angles corresponding to the covered portion. In practice, such reflectors are designed to provide coverage over a 90-degree cone. For a wider angle design, the reflecting cover will partially block the incoming wave at extreme angles, thus reducing the effectiveness of the device. The value of σ which holds over the operating region is based upon the projected area of the sphere and is given by

$$\sigma = \frac{4\pi}{\lambda^2} \ (\text{projected area})^2 \ .$$

In practice, a dielectric material having the continuously varying properties specified above is difficult to realize and some approximation must be made. One commercially available version of such a design, which provides excellent performance, consists of a spherical molded core and nine molded concentric shells. The core and shells are dimensioned in accordance with the theoretical Luneberg lens formula, but in discrete steps. The dielectric constant of the core is 2.0, and the shells have dielectric constants ranging from 1.9 to 1.1 in 0.1 steps. The material is a light-weight low-loss, plastic foam capable of being molded to a very accurate tolerance, both dimensionally and with respect to the dielectric constant.

19.4.3 The Van Atta Array

An interesting type of reflector, bearing the name of its originator, consists of an array of radiating elements which are interconnected in such a

fashion that the phase distribution corresponding to a received wavefront is transformed into that required to form a transmitting beam in the same direction. The transformation is accomplished by connecting each element to its diagonally opposed mate (with reference to the array center), all connecting lines having equal length.

The measured properties of such a device show that it provides angular coverage superior to that of a corner reflector. For example, a 4 by 4 dipole array with half-wave spacing produced a half-power return at an angle of 30 degrees, as compared with the 20-degree angle previously mentioned for the corner reflector. An outstanding advantage of the Van Atta Array is that it is essentially two dimensional, whereas the corner reflector requires substantial depth. This is of course an important consideration for airborne decoy applications.

19.5 Applications of Structure and Terrain Contouring

As previously indicated, the alteration of radar returns from structures and surrounding terrain may be accomplished through a combination of active and passive means. The active devices consist of electronic transponders, to be discussed in Section 19.6, which serve to augment normal radar returns. The passive devices can either augment or diminish the normal returns. In the first case, reflectors are used; absorbing materials accomplish the second objective.

Reflectors are suitable for the simulation of small cities, bridges, etc. Bodies of water such as lakes and rivers, which normally appear dark on the radar scope, may be concealed through the use of reflectors, thus removing important navigational aids. The design and distribution of the reflectors depends on the magnitude of the desired result and the viewing angles over which the effect is to be maintained. Returns beyond a certain magnitude cannot be produced satisfactorily through the use of reflectors, and more powerful techniques are required at this point. Electronic transponders or active target simulators fulfill the power requirements for the simulation of large cities. Many additional refinements in deception are available through the design of active circuitry of this type; for example, moving returns can be produced which will introduce errors in wind information obtained from them, thus compounding the difficulties of the bombing and navigation system.

Insofar as the reduction of radar returns is concerned, two methods are available. Structures such as buildings and bridges may be covered with various types of absorbent material in the first instance. Secondly, a con-

siderable amount may be accomplished by contouring fences or sloping land around buildings to destroy perpendicular relationships between walls and/or fences and the ground.

Each situation may require a particular combination of active and passive techniques.

19.6 Typical Electronic Target Simulator

The technique involved here is related closely to the problems of target masking and modification and is discussed for the sake of completeness. The equipment which performs this function, however, belongs to the general class of radar repeaters, and this subject is treated at length in Chapter 15.

A target simulator intended for ground use will differ from one employed to protect point targets such as ships and aircraft by the length of the returned pulse. The ground, being an extended target, will produce an echo of greater duration, and this characteristic must be simulated by the apparatus in order to provide adequate performance. It will also be observed that the range relationship used for the point target situation, i.e.,

$$\text{Range} \propto \sigma^{1/4},$$

does not hold in this case since, for a given antenna beamwidth, the greater the range, the greater the illuminated area becomes. In this connection, σ is defined on a unit area basis and the total effective echoing area is therefore proportional to the square of the altitude of the viewing radar. Thus the range relationship becomes

$$\text{Range} \propto \sigma_0^{1/2},$$

where σ_0 is the unit area value of ground reflection. This equation is the one which must be considered in establishing the power level of the target simulator.

The simplest physical realization of such a piece of equipment consists of a broadband receiver, pulse stretcher, and broadband transmitter. In a typical microwave application, traveling wave tubes might be employed in conjunction with a crystal detector to achieve a high, gain-bandwidth product. To avoid regeneration, the output pulse should be time-limited to less than the interpulse period of the received signal. In operation, the receiver is gated off while the pulse reply is being transmitted.

19.7 Experimental Examples of Target Alteration

Figures 19-2, 19-3, and 19-4 are airborne radar scope presentations which illustrate the alteration of targets through the use of corner reflectors. Five types of corners were used in obtaining this data:

<div align="center">

4 foot Triangular

4 foot Square

6 foot Square

6 foot Triangular

8 foot Triangular

</div>

In the case of the square reflector, the dimension refers to the edge, and in the triangular case, the indicated dimension is the altitude of the aperture.

Figure 19-2(a, b, and c) shows the effects produced at Oneida Lake, N. Y. In Figure 19-2a, this appears on the left, the upper boundary being the east end of the lake. The small peninsula seen at this end was formed with 10 corner reflectors (4 foot Triangular) placed on life rafts. In Figure 19-2b, this arrangement was disturbed by high surface winds which broke the anchoring. The individual corner reflectors can be seen as a series of battleships forming a flotilla across the lake. Figure 19-2c shows the same 10 corner reflectors laid out to form an island in the middle of the lake.
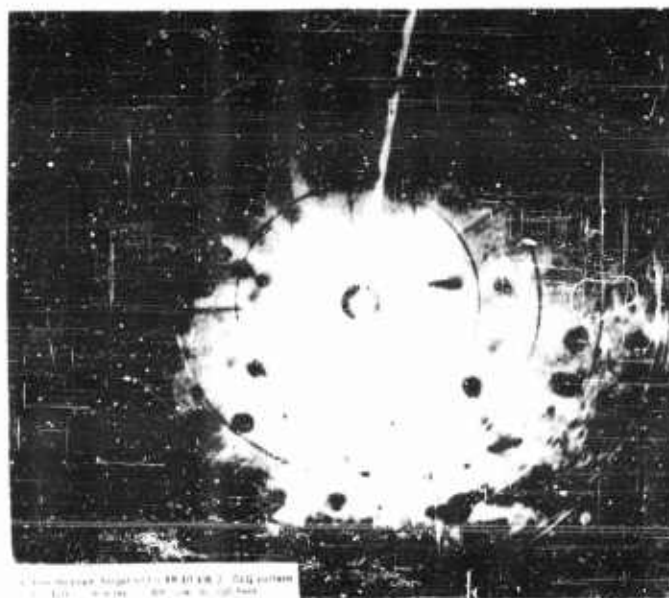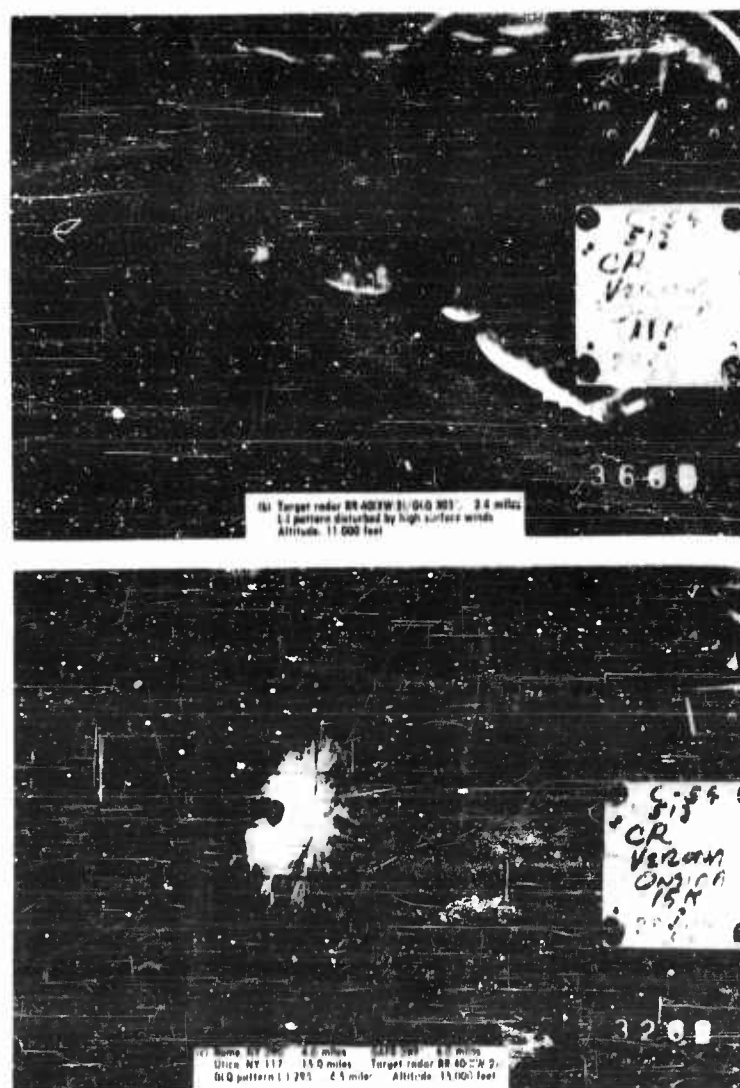
FIGURE 19-2   Corner Reflectors on Oneida Lake, N. Y.
(a) Peninsula formed with corner reflectors
(b) Surface wind effects (c) Island formed with corner reflectors

Figure 19-3 (a, b, and c) was taken in the region of Holloman AFB, New Mexico. Figure 19-3a shows the effect of 16 reflectors placed to produce a duplicate of Holloman AFB and a comparison of it with Holloman and Alamogordo. Figure 19-3b shows the effect of 20 reflectors set up to duplicate Holloman AFB together with the radar return from Holloman itself. Figure 19-3c shows the same situation at closer range.

Figure 19-4 (a, b, c, and d) was taken at Eglin AFB, Florida. The reflector complex, which faced north, consisted of thirteen 4 foot square, seventeen 6 foot square, seventeen 6 foot triangular, and twenty-one 8 foot triangular corners, each having a tilt angle of 28 degrees. Figure 19-4a shows the reflectors and the Eglin main complex at a distance of 174 miles at 180 and 185 degrees respectively. Figure 19-4b shows the reflectors at 94 miles and Eglin at 97 miles, at 176 and 184 degrees respectively. In Figure 19-4c, the reflectors are at a distance of 39 miles and a bearing of 174 degrees, with the Eglin main complex at 42 miles and 190 degrees. The final picture, Figure 19-4d shows the reflectors at 14 miles and 165 degrees, and Eglin at 17 miles and 205 degrees.

(b) Aircraft B50A   Altitude 25,000 feet   Radar APS 23
Reflector pattern   45 miles at 322   90 Reflectors
NAFB   50 miles at 331   6 four ft square
                                                6 six ft square
                                                6 six ft triangle
                                                6 eight ft triangle

29 April 1955

(c) Aircraft B50A   Altitude 25,000 feet   Radar APS 23
Reflector pattern   25 miles at 334   90 Reflectors
NAFB   29 miles at 348   6 four ft square
                                                6 six ft square
                                                6 six ft triangle
                                                6 eight ft triangle

21 April 1955
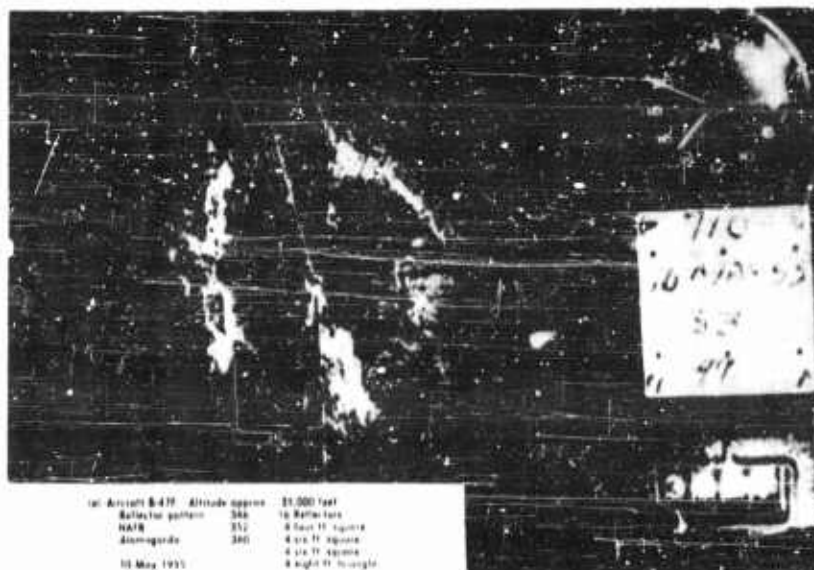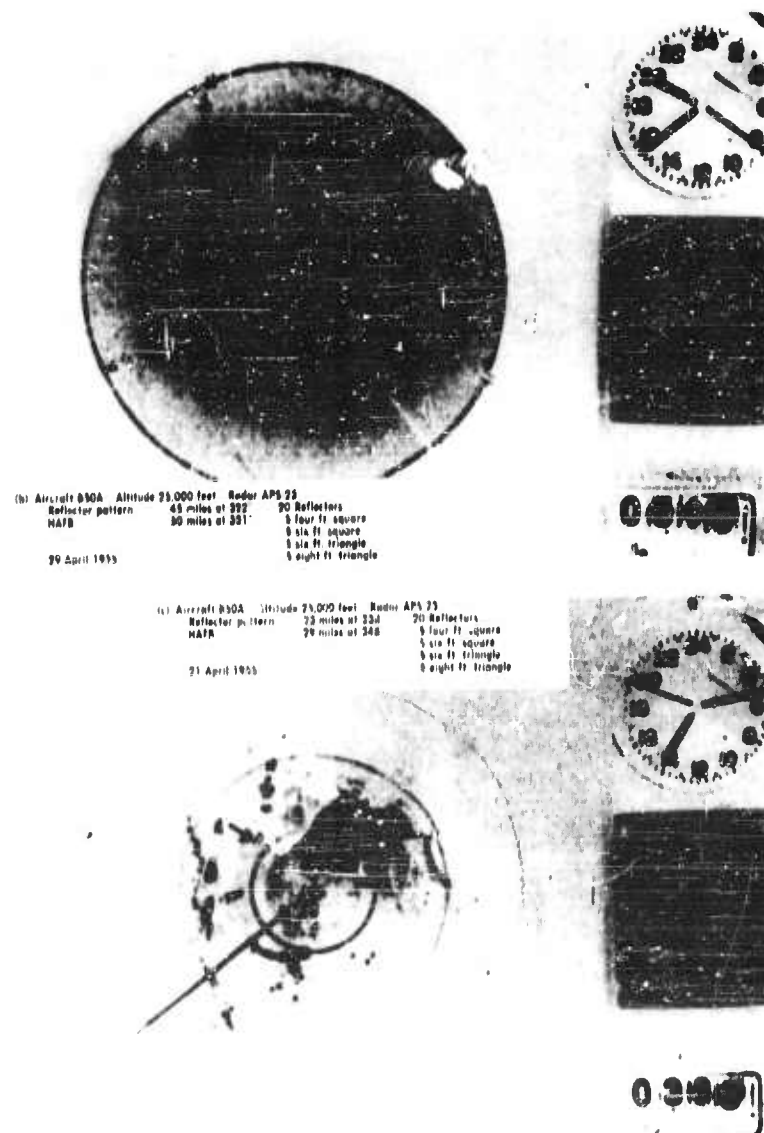
FIGURE 19-3   Reflectors Disposed to Produce Duplicate of Holloman AFB

(a) Reflectors, 174 miles, 180 degrees; Eglin main base, 174 miles, 135 degrees
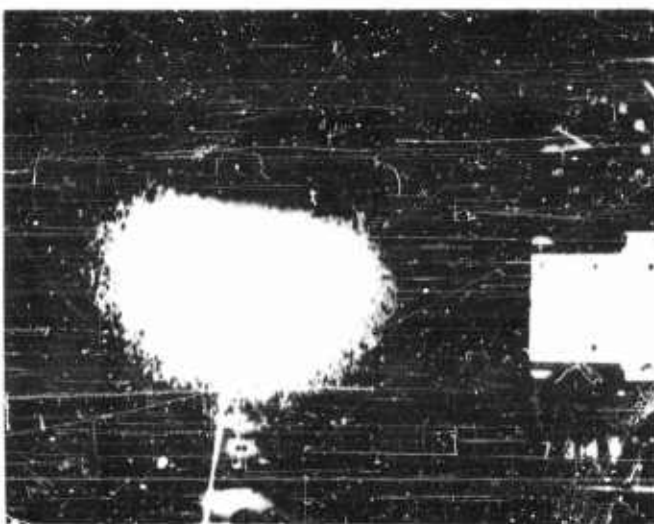
(b) Reflectors, 94 miles, 176 degrees; Eglin main base, 97 miles, 184 degrees
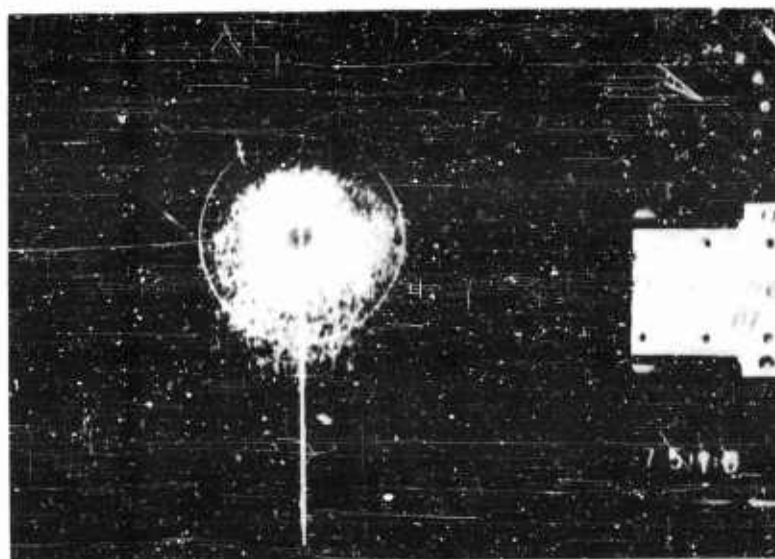
(c) Reflectors, 39 miles, 174 degrees; Eglin main base, 42 miles, 190 degrees
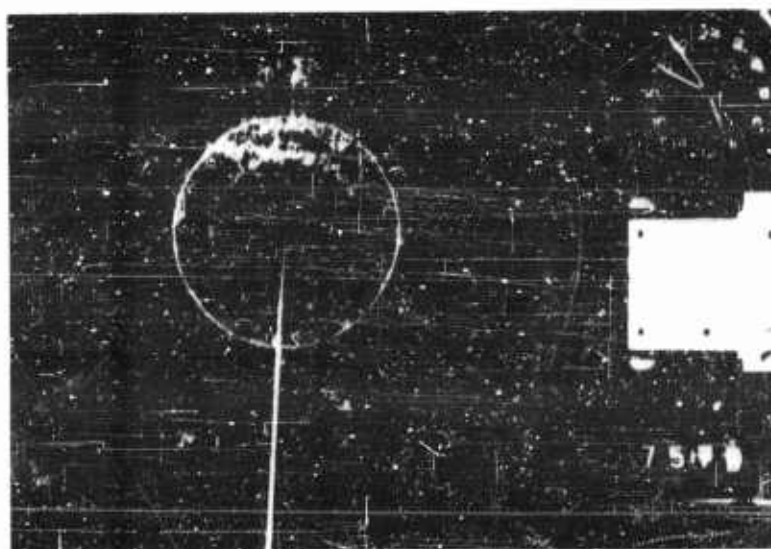
(d) Reflectors, 14 miles, 165 degrees; Eglin main base, 17 miles, 205 degrees
FIGURE 19-4 Reflector Complex, Eglin AFB, Florida

BIBLIOGRAPHY

Barasch, M. L., H. Weil, and T. A. Kaplan, Studies in Radar Cross-Sections-X Scattering of Electromagnetic Waves by Spheres, University of Michigan Report, prepared under RADC Contr. AF 30(602)-1070, dtd July 1956, and the bibliography contained therein.

Bedrosian, E., Curved Passive Reflector, *Electronics*, Vol. 29, p. 206+, December 1956.

Cohen, M. H. and R. C. Fisher, A Dual-Standard for Radar Echo Measurement, *IRE Transactions on Antennas and Propagation*, Vol. AP-3, No. 3, pp. 108-110, July 1955.

Court, G. W. G., Determination of Reflection Coefficient of Sea, for Radar-Coverage Calculation, by Optical Analogy Method, *Proc. IRE*, No. 6, pp. 827-830, November 1955.

Falkenbach, G. J. and R. J. Harrison, Research on Methods of Reducing Radar Cross Section of Aircraft, Battelle Memorial Institute Report, prepared under AFCRC Contr. AF 19(604)-1414, dtd 31 Aug 1957, and the bibliography contained therein.

Katzin, M., On the Mechanism of Radar Sea Clutter, *Proc. IRE*, Vol. 45, No. 1, pp. 44-54, January 1957.

Laasonen P., Determining Reflector Surface of Radar Antenna with Point Source Feed, *IRE Transactions on Antennas and Propagation*, Vol. AP-3, No. 4, pp. 180-184, October 1955.

Moses, L. F., "Investigation of Properties of Deutsche Magnesit S-28 Absorbing Material in the Microwave Region," ARDC Technical Report 57-101, Rome Air Development Center AD-131127, July, 1957.

Perov, G. I., The Problem of Calculation of the Effective Area of Reflection of a Surface Target in the Microwave Region, *Radiotehnika*, Vol. 11, No. 7, pp. 57-59, (In Russian), 1956.

Ramsay, J. F., Universal Scanning Curve for Wide Angle Mirrors and Lenses, *Marconi Rev.*, Vol. 19, No. 123, pp. 150-159, 1956.

Ramsey, J. F. and J. A. C. Jackson, Wide-Angle Beaming Performance of Mirror Aerials, Ibid., pp 119-140.

20

Decoys

M. E. BRODWIN

20.1 Introduction

One of the most effective counter measures against enemy radar defense systems is the technique of presenting the enemy with a large number of targets which completely saturate the defense system. When the defense is saturated, the survival probability of the attacking force increases very rapidly. A promising countermeasure which achieves this goal is the radar decoy. The premise is that if the defending radars are unable to distinguish between lethal vehicles and decoys then interceptors and missiles must be committed to both threats, thereby diluting and saturating the defense (References 1 and 2).

In this chapter, we first consider possible models of the enemy electromagnetic defense environment; early warning, ground-controlled interception (GCI), and missile control radars (References 3, 4, 5, 6, and 7). We then examine the techniques which the enemy may employ to distinguish between bombers and decoys. The simplest techniques are modifications of existing radar equipment. For example, in order to determine the relative size of targets it is possible to use the break-in range and subsequent information on aspect angle to distinguish between large and small targets. More sophisticated counter techniques require elaborate modifications, and, in some cases, new radar systems. For example, a defense might employ the polarization and statistical properties of the

echo, simultaneous comparison of echoes at different frequencies, and bistatic radars to identify the targets. Another possible technique is the distinctive infrared radiation of the different vehicles. Thus a study of possible techniques for identification is a necessary preliminary to the study of the simulation techniques.

Having examined techniques for discriminating between vehicles, we next consider the application of passive devices to simulate the echo of the bomber. An examination of the simulation requirements indicates that corner reflectors are capable of simulating the bombers for most angles at microwave frequencies. At the broadside angle, the target echoes are very large, and the simulation is more difficult. Since the simulation of broadside echoes places a severe weight and size penalty upon decoy equipment, we discuss, in detail, the problems involved in employing the large broadside echo for identification.

For frequencies below about 1,000 mc, the corner reflectors become too large to be carried on the vehicle. In this region, 50-1000 mc, it is necessary to employ electronic means to simulate the bomber.

The most useful active system for simulation is the straight through repeater. This device receives, amplifies and transmits an enhanced signal without altering any of the r-f characteristics. Unfortunately, the straight through repeater, under certain conditions at low frequencies, will oscillate. For those frequency bands where the straight through repeater fails, a gated repeater can be used.

Transponders are also considered for this application. The swept oscillator transponder, usually considered for false target generation, is analyzed to determine how well it would function for simulation. Another device, the pulse barrage transponder, which employs a wideband noise source to produce target echoes, is analyzed and experimental results are persented.

The active devices considered in this chapter differ, somewhat from repeaters in Chapters 15 and 19. Since the problem is to place sufficient energy in the main beam of the radar to simulate the bomber, relatively little power is required for effective deception. Consequently, techniques which may be feasible for decoys may not be useful for generating false targets and jamming.

Finally, we examine some of the problems encountered in designing equipment for simulating infrared radiation.

20.2 Characteristics of Decoy Missiles

Before discussing the defense environment in detail, it is helpful to note the significant features of a possible decoy. Tentative characteristics of a long range decoy are given below:

Range:	4000-5500 nautical miles
Speed:	Mach 0.8-0.9
Wing Span:	24 feet
Over-all Length:	33 feet
Gross Weight:	7500 pounds
Payload:	500 pounds
Altitude:	40,000-50,000 feet
Guidance:	Autopilot stabilization system using low drift rate, integrating rate gyro for directional control with provisions for preflight programming of turns in either direction.

A drawing of the decoy is shown in Figure 20-1. The forward section of the vehicle contains an array of corner reflectors for microwave simulation. Additional simulation for the broadside angles is provided by metal sheets imbedded in the plastic rudder. The payload is contained in wing sections close to the fuselage. The guidance equipment is located in the nose of the vehicle. Glass fiber construction is used for the wings, the forward section containing the reflectors, and the rudder.
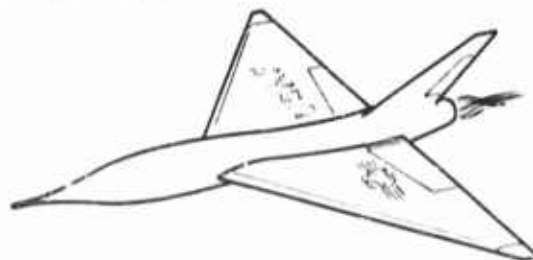
Figure 20-1   Drawing of a long range decoy

The purpose of the long range decoy is to saturate the interceptor and long range missile defenses along the attack corridor. A short range decoy can be used to saturate the inner defenses or short range, high firing rate, missile sites. Several can be launched from a bomber. It has been estimated (Reference 1) that three to five missiles per bomber are required for adequate dilution of enemy interceptors along the attack corridor.

The long range decoys can be employed in a number of ways. For example, they can be sent out in advance of the main bomber fleet. This tactic forces the defense to commit a portion of the fighters to intercepting the decoys, thereby reducing the force available for use against the main bomber fleet. Another possibility is to use the decoys for feint raids to draw off interceptors from the attack corridor. Decoys may also be employed for

seeding the corridor with chaff. Finally, the decoys can be used to harass
the defense by periodic fights independent of planned raids. In this manner,
the defense will dissipate some of its potential in destroying the decoy air-
craft.

20.3 Defense Discrimination Capabilities

The simulation requirements of the decoy are strongly dependent upon
techniques for discriminating between bombers, decoys and other long range
air breathing missiles. There are many possible ways for the defense to
discriminate between bombers and decoys, and it is the purpose of this sec-
tion to perform a preliminary evaluation of discrimination methods. In this
way, emphasis can be placed on simulation techniques which would be useful
against expected defense discrimination methods.

To aid the evaluation of different counter techniques, four assumptions will
be made.

1. The decoy is equipped with corner reflectors to increase the echo at
microwave frequencies and carries repeaters to simulate the bomber echo at
low frequencies.

2. The enemy will rely primarily on pulse radars in the early warning and
GCI operation.

3. The aerodynamic performance of bombers and decoys will be suffi-
ciently alike so that no identification between the two could be made by
means of air speed, altitude, or heading information.

4. Discrimination between bombers and decoys by airborne devices will
be of limited importance.

Target discrimination can be accomplished by both active and passive
methods. The passive methods include infrared detection and the intercep-
tion of aircraft electromagnetic radiation. The active methods involve modi-
fications of radar systems.

An infrared (IR) detector can be employed for target discrimination by
detecting differences in the strength of the radiation received from different
targets. The IR power will vary with the range and the aspect or viewing
angle of the target and this information must be available for analysis. The
principal disadvantage of ground based IR equipment is the limited range
resulting from atmospheric attenuation. Another serious disadvantage is
the requirement of clear skies.

A more promising technique is to sense the electromagnetic radiations
which are normally emitted by the bomber. In general, the bomber will be
emitting radiations which the decoys will not. Bombers will have active
electronic equipment; navigation and bombing radars, and jammers. With
regard to the former, the effectiveness of the passive detectors depends on

the duty ratio of the radars and prior information concerning the frequency
of transmission. It appears feasible to reduce the duty ratio, with the use
of storage tube techniques, to a value sufficiently small that the probability
of detection would be too low to be useful. With regard to jammers, it
would be necessary to mount jammers on some of the decoys. In fact,
decoys have been considered as a utility vehicle for jammers in raids where
mutual protection is feasible.

The most useful technique for passive discrimination relies on the dis-
tinctive signals which the decoy must radiate for low frequency simulation
(50-1000 mc). It will be shown that the simulation in the 50-1000 mc
region must be carried out by active means. Presently most of the proposed
simulation schemes involve some form of time sharing repeater. These sig-
nals appear as a normal return on an ordinary radar display. With special
equipment, the time sharing characteristics can be sensed and the decoys
identified. However, time sharing repeaters can also be mounted on the
bombers and the identification process becomes much more difficult.

In general, there is no high confidence technique for discriminating be-
tween bombers and decoys by passive means.

Discriminating between bombers and decoys by active detection refers
to a detailed examination of the characteristics of the radar echoes from
both vehicles. At present, the trend in radar design is to suppress detailed
information contained in the received pulses. It may be possible to dis-
tinguish between the two vehicles on the basis of differences in the average
amplitude of the echoes. Without simulation equipment, the larger aircraft
present a larger average radar cross section. This difference appears as a
difference in break-in range in present radars. With simulation equipment,
the differences in average amplitude are obscured by the noise in the radar
and the amplitude noise of the return. In addition there are variations in
echo power caused by changes in aspect angle. For most aspect angles a
small number of corner reflectors can simulate the target echoes. At the
broadside angles, the echo of the bomber is so large that the broadside echo
cannot be simulated by this configuration. It will be shown, that as the
vehicles pass through the defense network, the bomber presents a broad-
side aspect for a considerable portion of the flight. Consequently, it may be
possible to use the large broadside echo as a basis for discrimination. Tech-
niques for simulating the broadside echo will be discussed in a later section.

For other aspect angles, the decoy could appear larger than the bomber.
In general, considerable *a priori* knowledge is required about the average
cross section as a function of aspect angle. A further requirement is knowl-
edge of the aspect angle which can only be obtained from the track of the
vehicle and assumptions on the crab angle.

Another possible technique is to identify on the basis of the fluctuations in the target echo. The difference in the number of scatterers may appear as a difference in the fluctuations of the target returns. There are a small number of scatterers on the decoy in comparison with the assumed large number on the bomber.

If the decoys are equipped with simple corner reflectors, circularly polarized radars are capable of discriminating between the vehicles. The echo of a simple corner reflector is circularly polarized in a direction opposite to that of a normal target echo. This discrimination is obtained at a loss of 3 db in the effective radiated power. Techniques exist for "spoiling" the corner reflector to reflect both polarizations.

Another possible discrimination method is based on the variation of radar cross section with frequency. As the radar frequency is varied, the average scattering cross section at a specific aspect angle will vary. This technique would employ frequencies which are close to some resonant length of both vehicles. Thus, the returns at two different frequencies can be compared to determine the relative size of the vehicles.

In conclusion, the most likely methods for target identification can be listed in order of relative importance. The order represents a consideration of both the confidence placed in the identification technique and the extent of the modification of the radar defenses. For example, passive detection of active simulation equipment can be easily countered by placing simulators on the bombers.

1. Detection and analysis of unique low frequency simulated echoes.

2. Analysis of the polarization response of both vehicles.

3. Detection of jammer radiations from the bomber.

4. Determination of the average echo amplitude as a function of aspect angle.

5. Examination of differences in the statistics of target echoes.

6. Analysis of infrared radiations by airborne detectors.

7. Examination of the r-f frequency dependence of the target echoes.

Similarly an order of importance can be established for the simulation devices to be carried by the decoy. The order is based on consideration of the probable counter technique and the weight and size penalties imposed on the decoy. For example, consider the problem of simulating the large broadside echo. It will be shown that the use of the broadside echo requires significant modifications of the enemy defense system. In addition, the size penalty on the decoy imposed by large corner reflectors is a major modification of the decoy. Thus, the simulation of broadside echoes rates fairly low on the list.

1. Simulation of the average radar cross section of the bomber excluding the broadside angles
2. Simulation of the polarization response of the bomber
3. Simulation of jammers
4. Simulation of infrared radiation
5. Simulation of broadside echoes

20.4 Passive Simulation of the Radar Cross Section

The most important problem in simulation is to increase the scattering cross section of the decoy to approximate the scattering cross section of the bomber. Passive techniques for accomplishing this are considered in this section. We first discuss the application of tuned elements on the decoy fuselage for low frequency simulation in the 50-1000 mc range. Then we examine the application of corner reflectors to the high frequency range, 1000-10,000 mc. In both cases, the study is limited to the forward and rear aspects. The very high back scattering cross section observed at the broadside aspect of the bomber is discussed as a separate topic. Finally, we discuss the application of chaff to the decoy problem.

For a low frequency simulation, a possible technique is to resonate a part of the fuselage by electrically isolating a large section and using lumped tuning elements. For the analysis, the scattering cross section is approximated by a tuned wire antenna. The basic idea is that the scattering from a resonant element is much greater than that from a nonresonant element. The procedure is to compare the maximum scattering from an idealized tuned element with the scattering cross section of the bomber. Available experimental data, gathered by model range techniques, yield values of

TABLE 20-1. Scattering Cross Section of the B-47, B-52 Aircraft at Low Frequencies

Azimuth Degrees	B-47 Experimental 73 mc		B-47 Theoretical 66 mc		B-52 Theoretical 328 mc	
	max	min	max	min	max	min
0-10	65	5	5	5	5	3
10-20	65	5	5	5	3	1
20-30	15	5	10	5	3	1
30-40	65	5	26	10	1,800	1
40-50	195	5	16	10	10	2
50-60	15	5	15	10	10	2
60-70	5	9	30	15	40	3
70-80	195	5	100	30	180	9
80-90	575	65	1,000	100	25,000	50
90-100	900	50	1,000	120	25,000	3

scattering cross section as a function of aspect angle for the B-47 at a simu-
lated frequency of 73 mc (Reference 8). Theoretical values have also been
computed for a nearby frequency (Reference 9). Table 20-I shows the
extreme values of experimental and theoretical cross section σ for horizontal
polarization as observed in 10-degree intervals measured from the nose of the
aircraft.

The broadside and maximum scattering from reflecting wires is shown in
Table 20-II as a function of the length of the wire in wavelengths (Reference
10).

TABLE 20-II.   Scattering Cross Section of Wires

l/λ Wire Length	σ/λ^2 Broadside	σ/λ^2 Maximum	Angle of Maximum
0.50	0.99	0.99	0°
1.25	0.09	0.13	± 38°
1.50	0.87	1.70	± 45°
2.00	0.59	2.80	± 55°
2.50	1.53		
3.00	1.54		
3.50	2.80		
4.00	3.07		

Although these computations have been carried out for a length to dia-
meter ratio of 900, they yield order of magnitude results which are applicable
to the present case. For a half-wavelength dipole, which is the longest dimen-
sion of the decoy at 70 mc, the scattering cross section is about 18 square
meters. A possible arrangement would be to locate the wire along the edges
of the delta wing. This arrangement produces a horizontally polarized
response which is comparable to the values shown for the aircraft in the
forward direction. As the azimuth angle is increased, the scattering from the
dipole decreases while the bomber scattering increases. Thus, the simulation
is effective only over a restricted angle ahead of the decoy. As the frequency
increases, the decoy elements are longer, in wavelengths, and the scattering
decreases. For example, from Table 20-II, it is seen that the scattering from a
wire adjusted midway between the first and second resonance, $l/\lambda = 1.25$, is
about ten decibels less than the maximum of the half-wavelength dipole. The
design procedure would be to adjust the resonant element for best perform-
ance at the most important frequency at the low end of the frequency band
and accept the poorer response at the high end of the band.

Another possibility is the use of trailing wires in order to enhance the
broadside return. Referring to Table 20-I, it is seen that the broadside cross
section of the wire is orders of magnitude below the broadside scattering of
the bomber.

In conclusion, effective simulation by passive means at lower frequencies is limited by the basic fact that a scattering *linear* element cannot be made to appear arbitrarily larger than it actually is. Subject to the foregoing restrictions, it may be possible to make use of these techniques within a forward or rear sector of about ± 60 degrees. The enhanced scattering cross section is primarily horizontal and therefore subject to a relatively simple counter technique. In addition, the scattering exhibits an undes¹ ⁴ᵗ frequency dependence.

At ..igher freq⁴ .cies, 1-10 kmc, the scattering cross section ot the bombers is on the orde   to 100 square meters (Reference 9). The dimensions of the decoys have been chosen to accommodate corner reflectors which will produce adequate target echoes in this range.
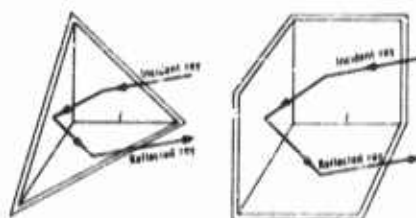
FIGURE 20-2   Corner reflectors

Corner reflectors consist of three mutually perpendicular conducting planes, Figure 20-2. The commonly used reflectors have circular or triangular sides. Another type of radar reflector that may be used is the Luneberg reflector (Reference 11 and 12).

In order to specify a corner reflector for the decoy simulation problem it is necessary to have the following information.

1. The relationship between the back scattering cross section and the dimensions of the reflector

2. The angular dependence of the back scattering cross section for different types of reflectors

3. The response of the reflector to the polarization o⁴ ⁴ incident field

4. The fluctuation in the reflected field

5. The relationship between back scattering cross section and frequency

The analysis of corner reflectors is based upon geometric optics (References 7, 13, 14, and 15).

For the case of a symmetrical triangular corner reflector, the maximum back scattering cross section is given by

$$\sigma_{max} = 4\pi l^4/3\lambda^2 \tag{20-1}$$

where σ_{max} = Maximum back scattering cross section; the scattering cross section for an incident ray making equal angles with the intersections of the planes forming the reflector

   l = Length of a side

   λ = Wavelength of radiation.

The corresponding relationship for a corner reflector with square sides is

$$\sigma_{max} = 12\pi l^4/\lambda^2 \qquad (20\text{-}2)$$

For a corner reflector with sides of circular segments, the scattering is given by (Reference 15)

$$\sigma_{max} = 15.59 l^4/\lambda^2 \qquad (20\text{-}3)$$

The scattering cross section of the circular reflector lies between the responses of the triangular and square reflectors. The choice of corner reflector depends upon the geometry of the vehicle, the required scattering cross section, and the angular dependence of the echo. For example, although the square corner reflector yields nine times the scattering of the triangular corner reflector, the lobe width of the square reflector is narrower than that of the triangular reflector.

The back scattering equations are based upon the assumption that the corner reflector can be represented by an equivalent uniformly illuminated aperture. This idea is valid when l/λ is much greater than unity. When the dimensions of the corner approach a wavelength, the scattering is no longer a monotonic function but exhibits variations in scattering cross section about some average value.

The Luneberg reflector has also been suggested for this application (Reference 12). This device consists of a plastic sphere with variable index of refraction, coated over a hemisphere with reflecting material. The incident rays focus on the reflecting surface and exit the reflector in the same direction. The maximum scattering cross section is given by

$$\sigma_{max} = 4\pi^3 l^4/\lambda^2 \qquad (20\text{-}4)$$

where l is the radius of the sphere. Although the lobe width of the Luneberg reflector is significantly greater than the circular corner, it occupies too much volume and is not usable for storing fuel during the early portions of the flight. In view of this, the Luneberg reflector is unsuitable for this application.

The angular dependence of the back scattering is calculated from the "lobe width" of the reflector. The lobe width is defined as follows: Consider a corner reflector illuminated by a monostatic radar. Along the symmetrical axis, the measured cross section is a maximum. Rotate the reflector until the cross section is one-half the maximum value. Twice the angle of rotation is defined as the lobe width of the reflector.

For a triangular corner, the lobe width can be calculated from (Reference 13)

$$\sigma/\sigma_{max} = (1 - 0.00076\ \Phi^2)^2 \quad \text{for} \quad \Phi \leqq 30 \qquad (20\text{-}5)$$

For a square corner, the corresponding equation is

$$\sigma/\sigma_{max} = (1 - 0.0274\ \Phi)^2 \quad \text{for} \quad \Phi \leqq 30 \qquad (20\text{-}6)$$

A graph of σ/σ_{max} as a function of Φ is shown in Figure 20-3. The circular corner reflector exhibits an angular dependence somewhere between these two extremes. For the triangular corner, the lobe width is 42 degrees whereas the square corner lobe width is 20 degrees.
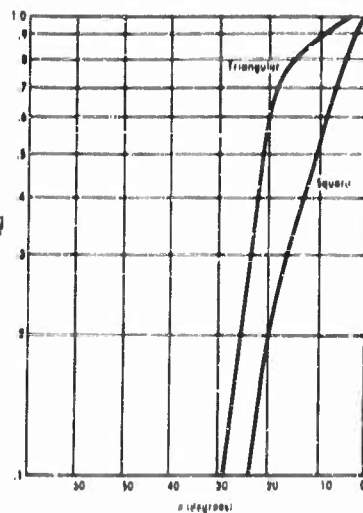
FIGURE 20-3 Scattering cross section versus deviation from symmetrical axis

The polarization of the reflected field is determined from the boundary conditions of a perfect conductor. For the case of triple reflection, the polarization response of the corner is the same as the response of a flat conducting sheet. Since the response of the corner reflector to circular polarization is well defined in contrast to the scattering from the bomber, a proper choice of radar receiver polarization will eliminate the decoy echo. A possible way to increase the cross polarized component is to replace a portion on one wall with a wire grid (Reference 1). Another possibility is to replace a wall of the reflector with a dielectric sheet (Reference 7). Since there is no phase reversal from the air-dielectric interface, the received polarization will not exhibit a reversal of the sense of polarization of the circularly polarized wave.

A single corner reflector exhibits very little fluctuation owing to the broad, smooth reflecting characteristics. However, fluctuation may be introduced in a number of ways. The most promising is to utilize the effects of a number of corner reflectors mounted on the vehicle. For 10 reflectors spaced 3 feet apart and operating at S-band, the total angular width between the first zeros on each side of the principal maximum is 0.16 degrees. This fine structure combined with the expected small random motion of the vehicle may produce an observed fluctuation comparable to the bomber fluctuation.

The frequency dependence of the scattering cross section can be determined from inspection of Eq (20-1) and (20-2). The response of the re-

fiector varies as the square of the frequency. Thus, the response at X-band will be nine times the response at S-band. The difference could be employed as a basis for discriminating between the vehicles. A promising way to "flatten" the response of the corner reflector is to introduce errors into the angles of the corners which compensates for the rising frequency characteristic (Reference 13). If the angles differ from 90 degrees, the response of the corner reflector is reduced. Let $\triangle\lambda$ represent the error in wavelengths of each edge of the corner reflector. The reduction in scattering cross section is shown in Table 20-III.

TABLE 20-III. Scattering Cross Section as a Function of Error Angle

			Decibel Reduction		
			1 db	3 db	10 db
Square Corner	$\triangle$	$=$	0.14λ	0.24λ	0.44λ
Triangular Corner	$\triangle$	$=$	0.20λ	0.35λ	0.62λ

Thus as λ is decreased, the error in wavelengths is increased, and the response is decreased. This result is shown in Figure 20-4 for two error angles. The curve with the 3-degree error was plotted by choosing a 1-db reduction at 2,000 mc. The 1.9-degree error curve was plotted by choosing a 10-db reduction at 10,000 mc.
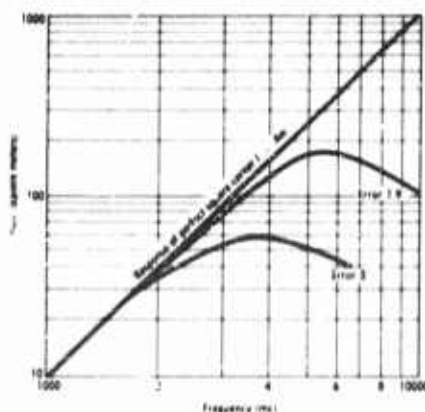
FIGURE. 20-4   Broadband response versus frequency

20.5 Broadside Echoes

The foregoing discussion excluded the large broad side echoes. It has been suggested that a possible way of distinguishing between decoys and bombers is on the basis of their echoes.

Information on the broadside returns of B-47 aircraft (References 8 and 9) indicate maximum cross-sections from 1.5×10^3 to 5×10^5 square meters. The lobe width, $\triangle\gamma$, is arbitrarily defined as the angle over which the scattering exceeds 1×10^3 square meters. The average lobe width (Reference 9) is on the order of 10 degrees.

To obtain an estimate of the importance of observation by the defense of the broadside aspect, the following assumptions are made.

1. The aircraft and radars are in the same plane.

2. The radars are arranged in an infinite rectangular lattice with spacing S_0 between adjacent radars.

3. The aircraft are flying a straight line path chosen at random through the radar network.
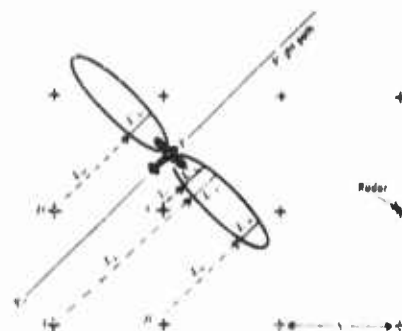
FIGURE 20-5   Geometry of broadside echo analysis

The geometry of the problem is shown in Figure 20-5. As the aircraft flies through the lattice, the broadside lobe "covers" different radar sites. When the radar is within the lobe, the aircraft is viewed at the broad ide angle. The time of observation of broadside aspects is proportional to the lengths of lines contained within the lobes after the aircraft have traversed a number of radars. For example, L'_A, L'_B, L'_C, and L'_D, are proportional to the observation time of the broadside angle by the radars A, B, C, D.

The total time of observation of all aspects is proportional to $L_A + L_B + L_C + L_D$. The fraction of the flight time in going from P_0 to P is

$$\text{Fractional Time of Observation} = \frac{L'_A + L'_B + L'_C + L'_D}{L_A + L_B + L_C + L_D} \quad (20\text{-}6)$$

For a small number of radars, this ratio depends upon the choice of flight path. A solution can be found for an infinite number of radars by mapping the entire lattice and lines into an elementary square, $ABCD$ (Reference 2, Part II).

The broadside lobes are mapped onto the radar sites at the corners of the square, Figure 20-6. The conclusion drawn from the analysis is given by

$$\text{Fractional Time of Observation} = \frac{\text{Area of Lobes Inside Square}}{\text{Area of Square}} \quad (20\text{-}7)$$
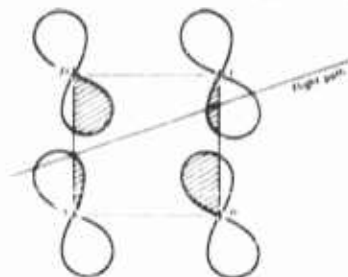
FIGURE 20-6   Broadside echo analysis

As the slope of the flight path is changed, the lobes rotate about the corners of the square, but the total lobe area contained within the square is constant. The lobe is assumed to be a sector of a circle of radius R and angular width $\triangle\gamma$. With these simplifying assumptions, the fractional time of observation is given by

$$\text{Fractional Time of Observation} = R^2 \, \triangle\gamma/S_0^2 \qquad (20\text{-}8)$$

Since the broadside echo area is very large, R is chosen as line-of-sight for an aircraft at an altitude of 40,000 feet. The results of this analysis are shown in Figure 20-7 which is a plot of Eq (20-8) for different values of

$\triangle\gamma$ as a function of radar spacing. To check the results of an infinite lattice analysis against a finite lattice, a geometric model was constructed, a random flight path chosen, and the line lengths measured. The measured value is also shown in Figure 20-7. For a corridor of length 1000 miles with radar spacings of 170 miles, the fractional time that the aircraft is observed at broadside angles is about 40 percent. The conclusion is that, it is possible in principle, to dis-

FIGURE 20-7   Fraction of flight time versus radar spacing

tinguish between the two vehicles on the basis of the broadside echo.

Another question which should be answered before considering the actual simulation devices is the minimum elevation angle required for tactically useful simulation. The minimum elevation angle is related to the fraction of the aircraft that pass within a specified range of the radar site. Consider a single radar with a maximum range represented by a circle of radius R_0. Assume that any straight track at a given slope is equally likely. If the spacing between parallel tracks is denoted by $\triangle S$, the number of tracks that pass within a distance S from the radar is given by

$$m = 2S/\triangle S \qquad (20\text{-}9)$$

The total number of tracks that pass through a radar with range R_0 is given by

$$N = R_0/\triangle S \qquad (20\text{-}10)$$

The fraction of the total number of tracks that pass within a distance S from the radar is then

$$m/N = S/R_0 \qquad (20\text{-}11)$$

For the multiple radar case the fraction increases. Assume a rectangular lattice of radars spaced R miles apart. A corridor, C miles wide, is chosen at random and directed toward the target. Let

$\triangle S$ = Spacing of tracks which are equally distributed throughout the corridor

s_k = A chosen radius

S_k = Radius of the maximum range of the kth radar

L_k = Overlap of S_k on the corridor

l_k = Overlap of s_k on the corridor

P = Total number of radars observing the corridor.

Then the number of aircraft that pass within a circle of radius s_k is $l_k/\triangle S$. The number within the surveillance range of the kth radar is $L_k/\triangle S$. The fraction of the total number of tracks that fall within circles of radius s_k is

$$m/N = \sum_1^P l_k \; \sum_1^P L_k \tag{20-12}$$

The answer was obtained by graphical construction and is shown in Figure 20-8.* For each assumed ground range and altitude, the elevation angle can

be determined. In this manner it is possible to obtain an estimate of the minimum elevation angle corresponding to a given percentage of aircraft which are permitted to be identified.

For an elevation angle of 10 to 15 degrees, about 15 percent of the attacking force will exhibit greater elevation angles and may be identified. Consequently, the simulation device is required to possess a vertical lobe width of about 10 to 15 degrees.
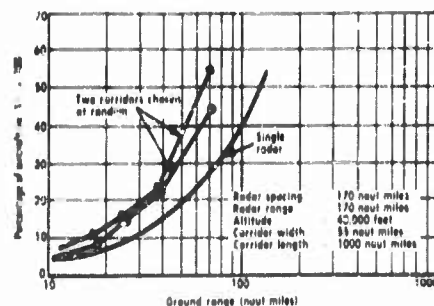
FIGURE 20-8 Percentage of aircraft which pass closer to the radar than the abscissa

The methods for increasing the broadside echoing area are both passive and active. Three passive methods for increasing the broadside area have been considered; triple corner reflectors, flat sheets, and double corner reflectors. The first technique is discarded due to the large size required to exhibit the large required scattering. The second method produces the required echo area but the lobe is much too narrow. The double corner appears to be the most promising passive technique. The maximum scattering cross section of the double corner is given by

*The construction leading to Figure 20-8 is of considerable use in jamming problems. It yields a relationship for the fraction of aircraft failing within a given breakthrough range.

$$\sigma_{max} \approx (8\pi/\lambda^2)(lL)^2 \qquad\qquad (20\text{-}13)$$

where l is the length of the corner and L is the length of a side. For L about 0.3 meters, and l about 7 meters, the scattering cross section is 1×10^8 square meters at X-band and 8×10^8 square meters at S-band. Thus a double corner fulfills the requirement for scattering cross section. The reflector must also present this cross section over the aspect angles comparable to the bomber. The double corner, with the long dimension parallel to the principal axis of the vehicle, exhibits a very large elevation lobe width and a small azimuthal lobe width. The elevation lobe width is approximately 60 degrees. The azimuthal lobe width in radians is given by

$$W = 0.91 \ (\lambda/l) \qquad\qquad (20\text{-}14)$$

For a double corner, seven meters in length, the azimuthal lobe width is about .08 degrees. This value is much too small to be useful. To obtain a 20-degree lobe width at S-band requires a length of about 0.23 meters. At X-band the lobe width would be about 7 degrees. A possible design would consist of an array of corner reflectors on each side of the aircraft.

The principal disadvantage of the double corners would be the increased space occupied by the reflectors. Although passive methods appear promising for simulating the broadside echoing area, the possibility of an active simulation technique has also been investigated. The active techniques are considered in Section 20.6.

Another technique for preventing the identification of decoys is to utilize the obscuring effect of chaff. Chaff may be particularly valuable for this purpose in the frequency spectrum below 1000 mc, where it may be difficult to confuse the radar identity of decoys and bombers. The effectiveness of chaff is analyzed for the specific case in which the decoy scattering is less than the bomber scattering.

The aspect of chaff sowing that is unique to the decoy problem occurs when the density of the chaff is so light that some of the bombers and some of the decoys may be observed. The amount of chaff necessary to prevent the enemy from detecting the aircraft can be calculated by assuming that the chaff return is indistinguishable from the receiver noise. The method is based upon the relationship between the blip to scan ratio and the signal to noise ratio, (References 16, 17, and 18). This relationship is shown in Figure 20-9. The assumed probability density for the input signal to noise ratio is

$$\frac{1}{X} \ e^{-x/X}$$
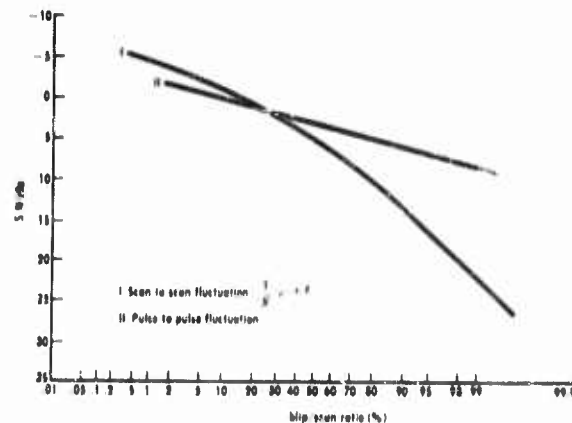
FIGURE 20-9   Blip scan ratio versus signal to noise for ten
hits on a fluctuating target

where X is the input signal to noise ratio, and $\overline{X}$ is the average of X over all
target fluctuations.

Since the return from chaff is added to the reflected signal from the air-
craft the presence of chaff decreases the signal to noise ratio. Thus, the sow-
ing of chaff to conceal an aircraft reduces the blip to scan ratio. For example,
if the signal to noise ratio is 20 db for the aircraft alone, the blip to scan
ratio would be 98.2 percent. Assuming that the echoing area of the decoy is
10 db less than the echoing area of the bomber, the signal to noise ratio of the
decoy would be 10 db. This corresponds to a blip to scan ratio of 80 percent.
Both targets would be detected at this ratio. If a moderate amount of chaff
is sown, the signal to noise ratio for both targets would deteriorate. Assuming
that the chaff reduces the signal to noise ratio of the bomber to 10 db, the
blip to scan ratio of the bomber becomes 80 percent. The same amount of
chaff reduces the blip to scan ratio of the decoy from 80 to 15 percent and
the decoy may not be detected. Thus, if the decoy echoing area is less than
that of the bomber, a few of the bombers may not be detected, but a larger
number of the decoys may also escape detection. Although a light sowing of
chaff reduces the dilution of the defenses by the decoys, some of the bombers
will not be observed. Thus the net effectiveness of the mission may still be
increased by the use of chaff.

20.6 Active Simulation of Airborne Targets

In the previous discussion it was concluded that the passive techniques
were adequate for the higher frequencies. The crossover point between high

and low frequencies is arbitrarily chosen at 1000 mc pending full scale model tests. In the frequency range between 50 and 1000 mc, the tuned fuselage techniques were shown to be adequate for forward aspects but too low for other aspects. What is required for simulation in the low frequency region is a device which would broadcast signals sufficiently similar to the bomber echo to prevent identification. If the transmitted signals were distinctly different from the skin echo, they would form the basis for discrimination. In this case, the device would be mounted on both the bomber and the decoy. In this section, it is assumed that the skin return of the bomber is about ten times the skin return of the decoy.

A number of active simulation devices have been studied for this problem. Among them are straight through repeaters, gated repeaters, swept oscillator transponders, and pulse barrage transponders. The ideal device would be an amplifier which would broadcast the received pulse at a higher amplitude without distinctive characteristics. A system which fulfills these conditions is a straight through repeater. When there is sufficient isolation between the receiving and transmitting antennas no oscillations take place. The criteria for sufficient isolation depend upon the characteristics of the amplifier, the geometry of the system, and the gain pattern of the antennas. To prevent large fluctuations in output power, it is desirable to keep the leakage signal small with respect to the received radar pulse. Another effect of excessive coupling is lengthening of the trailing edge of the radar pulse. This lengthening is on the order of the delay of the amplifier for small leakage. For larger coupling, the trailing edge may be lengthened by several times the delay of the amplifier.

The required gain of the repeater is calculated for the case in which the coupling between the antennas is neglected. For adequate simulation, the received power at the radar from the repeater is equal to the received power from the skin of the bomber. If K' is the gain of the system, between the two antennas, and G_r is the gain of the repeater antenna, the system gain is given by

$$K' = 4\pi\sigma/\lambda^2 G_r^2 \qquad (20\text{-}15)$$

The power output of the repeater consists of both the amplified energy from the radar, and the amplified energy due to antenna coupling. The power P_s delivered to the transmitting antenna, is given by

$$P_s = \frac{KG_r\lambda^2}{(4\pi)^2}\left(\frac{WG_t}{R^2} + \frac{P_sG_te^{j\Phi}}{a^2}\right) \qquad (20\text{-}16)$$

where W is the power transmitted by the radar, G_t is the maximum antenna gain of the radar, a is the separation between the repeater antennas, and Φ is an arbitrary phase angle depending upon system delay and antenna separation. From the previous condition for adequate simulation, the amplifier gain is given by

$$K = \frac{4\pi\sigma \, / \, G_r^2 \, \lambda^2}{1 + \sigma \, e^{j\Phi}/4\pi a^2} \qquad (20\text{-}17)$$

Another way of stating the simulation condition is that the beam power of the repeater is equal to the scattered power from the bomber, or

$$P_s G_r = W G_t \sigma / 4\pi R^2 \qquad (20\text{-}18)$$

The previous results hold only for the case of two isotropic antennas. For the nonisotropic case, let $G_r(\theta_t)$ be the antenna gain in the direction of the radar and $G_r(\theta_s)$ be the antenna gain along the line joining the two repeater antennas. The amplifier gain for this case is

$$K = \frac{1}{\lambda^2} \; \frac{4\pi\sigma/G_r^2(\theta_t)}{1 + (\sigma/4\pi a^2) \, |G_r^2(\theta_s)e^{j\Phi}/G_r^2(\theta_t)} \qquad (20\text{-}19)$$

The system gain K' is the ratio of the power delivered to the transmitting antenna, to the power at the input terminals of the receiver.

$$K' = \frac{K}{1 - |KG_r^2(\theta_s)\lambda^2 e^{j\Phi}/(4\pi a)^2|} \qquad (20\text{-}20)$$

Reasonable values for the amplifier gain can now be calculated. Since there is no control over the phase term, $e^{j\Phi}$, let this term be unity. To prevent oscillation,* let

$$\frac{K G_r^2(\theta_s) \, \lambda^2}{(4\pi a)^2} \; < 1 \qquad (20\text{-}21)$$

or

$$K\beta < 1$$

*The straight through repeater analysis differs from the usual feedback analysis where the factor $K\beta$ is chosen greater than unity to stabilize the system. In the repeater, Φ is a rapidly varying function of frequency and the only way to insure stability is to require that $K\beta$ be less than unity. If $K\beta < 1/3$, the maximum fluctuations of amplitude due to the phase term is 3 db.

where

$$\beta = G_r^2 (\theta_s) \lambda^2 / (4\pi a)^2$$

The maximum scattering cross section that can be simulated is calculated with the aid of Eq (20-21). Since the requirement of stability implies that $K \approx K'$, Eq (20-15) is substituted for K in Eq (20-21) and the 1/3 criterion is applied. Therefore, for stable operation,

$$\frac{\sigma}{4\pi a^2} \cdot \frac{G_r^2 (\theta_s)}{G_r^2 (\theta_1)} \leq \frac{1}{3} \qquad (20\text{-}22)$$

The equality sign is used to calculate the maximum scattering cross section which can be simulated. The maximum scattering cross section is shown in
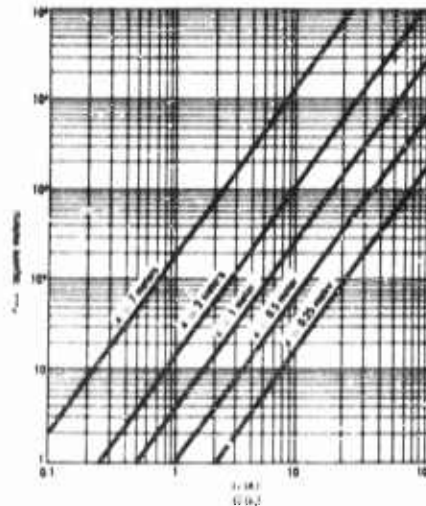
FIGURE 20-10   Maximum cross section ver-
sus relative antenna gain

Figure 20-10 as a function of the separation of the repeater antennas and the ratio of the gains.

The stability criterion depends upon the inverse relationship between the gain of the amplifier and the square of the wavelength assuming that the antenna gain is constant, Eq. (20-15). For short dipoles, the antenna gain is constant and the amplifier should be designed with an appropriate filter.

The stability criterion, Eq (20-22), also depends inversely on the spacing between the antennas, a, and the ratio of the antenna gains. Since the dimensions of the decoy are fixed, there will be some minimum frequency at which the repeater will oscillate.

For operation *below* the oscillation frequency, other active simulation devices are required. These consist of gated repeaters, swept oscillator transponders, and pulse barrage transponders.

A block diagram of a gated repeater is shown in Figure 20-11. The gating circuit supplies a rectangular gating pulse to the receiving amplifier in the sequence ON-OFF. A gating pulse is also applied to the transmitter in the sequence OFF-ON. The received energy, amplified during the ON time, charges the delay line. Assuming that the radar pulse is much longer than the gating period, the output of the delay line is a train of short pulses with

FIGURE 20-11   Block diagram of a gated repeater

an over-all length approximately equal to the radar pulse length. The delay is adjusted to one-half the repetition period of the gating pulse. Thus, during the receiver ON time, the delay line stores the received energy. During the receiver OFF time, corresponding to the transmitter ON time, the energy stored in the delay line is transmitted. If the delay time is greater or equal to the one-half of the gate cycle, and less than the total gate period, the receiver is never ON at the same time as the amplifier. Consequently, the system cannot oscillate. Thus, the gated repeater quantizes the radar pulse and transmits an enhanced version of the received pulse. The length of the transmitted pulse train can at most differ from the radar pulse length by a gate period. It will approach the radar pulse length as the gating period is reduced.

A preliminary step to the system design of the gated repeater is to determine how the radar receiver responds to the pulse train. It is necessary to determine the output pulse shape at the radar, the reduction in amplitude caused by the gating operation, and the factors affecting the choice of the gating period and duty cycle.

The output of the repeater can be expressed by

$$
\begin{aligned}
f(t) &= a \cos \omega_1 t && \text{for } T_1 < t < T_2,\ T_3 < t < T_4 \ldots T_{n-1} < t < T_n \\
&= 0 && \text{for } t < T_1,\ t < T_3 \ldots t > T_n
\end{aligned}
\tag{20-23}
$$

Equation (20-23) represents a series of pulses with a carrier frequency of ω_1. If ω_2 is the frequency of the local oscillator in the radar receiver, the input to the i-f filter is given by

$$
f(t) = Re \ \frac{a}{2} \left| \exp j(\omega_1 - \omega_2)t + \exp j(\omega_1 + \omega_2)t \right| \tag{20-24}
$$

An approximate result is obtained for a flat-topped rectangular filter with bandwidth B. For a repeater duty factor of $1/2$, this is given by

$$
F(t) \approx \frac{1}{\pi} \sum_{k=2p}^{2N-1} Si|x - k\Phi| - Si|x - (k+1)\Phi| \tag{20-25}
$$

where $x = \pi B t$ and $\Phi = \pi B T$, $N =$ number of pulses in the pulse train and $p = 0,1,2,3, \ldots$. T is the ON time of the pulse in the pulse train. Equation (20-25) represents the output pulse shape at the radar. For ex-

ample, consider a 5-μsec radar pulse with a radar receiver passband of 250 kc. In this case, Φ is $(5/4)\pi$, and the output pulse shape is shown in Figure 20-12. The response of the radar to five 0.5-μsec pulses is also shown. The shape of the response of the filter is approximately the same for both the radar pulse and the repeater simulated pulse. If the radar passband is increased beyond the optimum value of $(5/4)\pi$, the individual gating pulses are resolved. For example, Figure 20-13 shows the response when two 0.5-
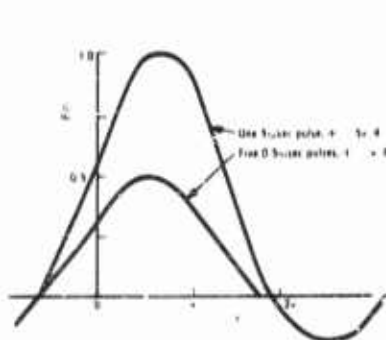
FIGURE 20-12 Output pulse shape for gated repeater (optimum bandwidth)
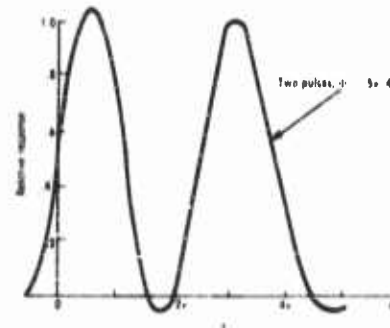
FIGURE 20-13 Output pulse shape for gated repeater (large bandwidth)

μsec pulses are applied to a receiver with a bandwidth of 2.5 mc. This result suggests an attractive countermeasure against the gated repeater: a wideband auxiliary receiver with a pulsewidth discriminator could be employed to detect the presence of gated repeater transmissions. Discrimination against a gated repeater is also possible on the basis of its distinctive pulse spectrum.

The next problem is to compute the reduction in amplitude caused by the gating of the radar pulse. For example, Figure 20-12 shows that the amplitude is reduced by a factor of 2 for a duty cycle of one-half. The problem can be analyzed for a range of duty cycles and pulsewidths by determining the energy in the pulse train as compared to the energy in the original radar pulse at the output terminals of the filter. The energy loss ratio is given by

$$E_2/E_1 \approx (3/8)NBT \qquad (20\text{-}26)$$

for the parameters in the region of $NBT = 1/2$. Equation (20-26) represents the loss introduced by the gating operation. This result was obtained by Fourier transform methods followed by graphical integration of the resultant energy expressions.

The choice of gate period does not effect the filter loss. If τ is the total length of the radar pulse, and D is the duty factor of the gate cycle,

$$NT = \tau D \tag{20-27}$$

As the gate period is changed, D is constant, and the product NT is constant for a specific radar. Thus BNT is unchanged and the energy loss is constant.

An inspection of the block diagram, Figure 20-11, suggests that a saving in weight and power can be gained by reducing the gate period to a small value. A shorter delay line would occupy less volume and introduce less loss into the system. However, a reduction in gate period increases the complexity of the gate generator. Thus there is some optimum gate period representing a compromise between the reduction in delay line losses and increased complexity of the gate generator.

Another factor that affects the energy loss and the design of the gated repeater is the loss introduced by smaller duty factors. Once again the design is a compromise. If the amplifier benefits from small duty cycle operation, then a reduction in duty cycle implies a reduction in weight of the power supplies. However, if D is reduced, NT is reduced, and the energy loss increases as NT.

The next problem that underlies the design of all gated repeaters is to determine the gain, maximum power output, and dynamic range of the amplifiers. The over-all system gain of the gated repeater is the same as the required gain of the straight through repeater modified by the gating loss. The maximum power of the repeater, at some minimum range R_m is given by Eq (20-18). For ranges less than the minimum, assume that the amplifiers are so limited that the output power of the repeater is too low for adequate simulation. Thus, for ranges less than the minimum, the defense can identify the vehicle as a decoy. The problem of minimum range can then be restated in terms of how many decoy vehicles are permitted to be identified without excessively reducing the effectiveness of the strike. An estimate of the minimum range can be obtained by means of Figure 20-8 which is a graph of the fraction of aircraft that pass closer to the radar than at a specified range. If 10 to 20 percent of the decoys can be permitted to be identified, the minimum range is about 25 miles. Assuming, for the low frequency region, an echo area of 50 square meters, a radar of 100 kw, with an antenna gain of 20 db, and a repeater antenna gain of unity, then a maximum power of .025 watts or 14 dbm is required of the repeater. This number is not, however, the output power from the transmitting amplifier. This value must be adjusted for the loss due to gating, and the antenna loss. A more reasonable estimate of the maximum output power of the transmit-

ting amplifier is about 0.2 watts. Low output power is characteristic of all simulation devices in contrast with jammers.

If the gated repeater is considered for simulating the broadside echo, the minimum required power can be calculated by means of Eq (20-18). For the same minimu range and an echo area of 1000 square meters, the maximum output power is 4 watts. This figure should be reduced by the gain of the directive antennas employed for simulating the broadside angle.

The dynamic range requirement is also not very severe. If it is desired that the repeater operate over a range of from 10 to 250 miles, the required dynamic range is about 27 db.

Since the decoy aircraft will fly either intermingled with the bombers or on a separate flight, their spacing will be on the order of 5 to 10 miles. It is therefore necessary to examine the possibility of a repeater interacting with an adjacent repeater and producing sustained oscillation between the two devices. The interaction range R_i is defined as the range at which the gain of the repeater is equal to the transmission losses. Oscillation will occur for any smaller range. The interaction range is calculated from the free space transmission formulas and it can be shown that the interaction range is given by

$$R_i = \sqrt{K' \ (G_r \lambda/4\pi)} \qquad (20\text{-}28)$$

for identical repeaters. The maximum value of K' occurs at the highest operating frequency. For a frequency of 1000 mc, an echo area of 50 square meters, and an antenna gain of unity, the system gain is 38.5 db. If an extreme case of antenna orientation is assumed, $G_r = 20$ db, the interaction range is 640 feet. This small range does not present any problems of interaction for the decoy repeater.

The minimum altitude at which the repeater can operate without oscillation is determined by assuming that the repeater is above a plane conducting earth and that there is an image repeater equidistant on the opposite side of the plane. The minimum altitude is one-half the interaction range, 320 feet.

The last problem to be discussed before examples of specific systems are described is the behavior of the repeater in the presence of expected signal densities. The repeater will ordinarily respond to all radiation within its frequency band. The amplifiers may overload or limit if the sum of the retransmitted signal power exceeds the capacity of the amplifier. It may thus be necessary to increase the power handling capacity of the amplifier. It is expected that the repeater will not be required to simulate more than two signals at the maximum power level, and consequently the increase in capacity may be only 3 decibels.

Another aspect of this problem is the fact that the repeater will amplify all low level signals and the sum of these low level signals may exceed the capacity of the amplifier. This effect may be reduced by noting that the extraneous signals are all lower than the main beam radiation of the radar. To reduce the response to the undesired traffic, a threshold device can be inserted at the output of the receiving amplifier. The threshold detector operates by turning on the transmitting amplifier whenever the received signal exceeds a specified level. Thus the system will only repeat the radiation from the main beam of the radar.

The previously discussed systems operate on the repeater principle. It is possible, however, to devise transponders which will also simulate the bomber. Two systems have been analyzed, the swept oscillator transponder and the pulse barrage transponder.

The principle of the swept oscillator transponder is to use the received pulse to trigger a sweep generator which rapidly tunes the oscillator throughout the frequency band. The swept oscillator produces an output pulse on the radar screen which resembles a large target. Since the allowable delay between reception and transmission is only a fraction of the pulsewidth, a fast tuning oscillator is required.

The decoy transponder is much simpler than systems employing voltage tunable tubes for false target operation or range gate deception, since programming is not required for decoy simulation.

Assuming that the sweeping action reduces the required power by a factor Q, the required output power P_j in the passband of the radar receiver is given by

$$P_j = Q\, P_t\, G_t\, \sigma/4\pi\, R^2\, G_j \qquad (20\text{-}29)$$

where P_t is the power output of the radar (100 kw), G_t is the gain of the radar antenna (20 db), σ is the scattering cross section of the bomber (50 square meters), R is the range (170 miles), and G_j is the gain of the repeater antenna (unity). Under these conditions, the power required of the transponder is about 0.05 watts for $Q = 1$

The degradation factor Q can be calculated from the relationship between the magnitude and duration of the output pulse at the defense radar, and the characteristics of the sweeping signal.

An analysis of the swept oscillator transponder demonstrates that it is a feasible means of simulating the bomber. The principal disadvantage is the ease with which the swept oscillator transponder could be countered. A simple technique would be to detune the radar receiver. All the pulses that appear are therefore decoys. Another possibility is to use an auxiliary receiver, tuned

to an adjacent frequency, to "blank" the decoy targets. The counter-counter-measure is to require the same transponder on both bombers and decoys.

The pulse barrage transponder operates by using the received radar pulse to trigger a broadband noise source. When a radar pulse of sufficient magnitude is received, a burst of noise is transmitted on all frequencies in the band. The integration of a PPI display masks the noisy characteristic of the signal. An experiment was undertaken to determine how much noise should be added to the skin echo of the decoy to confuse the operator. Two pulses were presented to an operator on a PPI display. The bomber pulse level was set 10 db greater than the decoy pulse. Both targets were presented to the operator at different locations, and the operator was asked to decide which target was the bomber. Noise bursts were then introduced into the system in coincidence with both target pulses. The noise bursts were adjusted to the same length as the target pulses. The noise power was increased until the identification error of the operator approached 50 percent. The resulting ratio of rms noise power to the peak bomber power was 20 db. The results of this simplified experiment were considered too high since the noise contribution of the receiver and the fluctuation of the target were ignored. Assuming a ratio of rms noise power to peak bomber power of N, the required output power P_t in the passband of the receiver, is given by Eq (20-29) with Q replaced by N. For the radar characteristics illustrating Eq. (20-29) and a receiver bandwidth of 250 kc, the required power density is 0.2 watts/mc. When the additional factors of receiver noise, target fluctuation, and aspect angle are considered, the required power density is probably less by at least an order of magnitude, 0.02 watts/mc.

Another problem associated with the use of transponders for decoy simulation, is the variation in range at which the transponder commences operating. The received power at the transponder is a function of the transponder antenna gain as well as the range from the radar. Since the threshold at which the transponder commences operation is preset, the range at which the decoy first appears upon the PPI will depend upon the aspect angle of the decoy. Compression circuits are required in the transponder to insure that the transponder commences operation at the desired range without too much variation.

Since the length of the output pulse of the transponder is fixed, it is therefore subject to simple counter techniques. Although variable pulse length circuits were considered, these circuits require extensive receiver and programmer developments. With fixed pulse length operation, transponders must be mounted on both bomber and decoy. Another major difficulty is the weight and power requirements for a unit covering the 50 to 300-mc band.

20.7 Infrared Simulation

Since the problem of identifying decoys and bombers by IR is uniquely

different from the electronic problem, this aspect of simulation is discussed separately.

The defense model consists of a large IR search system and surveillance radar carried in a high altitude aircraft which relays target information to the ground. The basis for identification is the amount of IR radiation produced in the wavelength band of the detector. The identification can be performed by measuring the IR radiation as a function of range and aspect angle.

The expected range at which the decoy could be detected is based upon the test results which indicate a 200-mile range for a fighter which radiates 120 watts per steradian at the tail aspect in the 2 to 3 micron region (Reference 2). This range is optimistic since atmospheric attenuation and sky noise were neglected.

A reasonable assumption concerning the magnitude of the radiation is that the IR radiation is roughly proportional to the fuel consumption for similar operating temperatures and efficiencies. Thus a bomber can be expected to radiate 25 times as much IR energy as the decoy. The range at which the decoy is first observed would then be about one third to one fifth the detection range of the bomber at the same aspect. The difference in the detection range or break-in range can be used as a basis for discrimination. As the vehicle approaches the detector, the small forward radiation limits the detection range to 30 to 40 miles for the bomber and 5 to 10 miles for the decoys. As targets pass the search plane, the rear aspects allow much larger detection ranges.

It is unlikely that an examination of the spectral distribution of the sources will yield useful information, since both engines will burn similar fuels at the same temperatures. Thus, the most attractive IR basis for discrimination is the amount of IR energy radiated by the vehicles.

If it is expected that the enemy will use IR detection devices for discrimination, it is possible to place a hot body on the decoy to approximate the IR radiated by the bomber. The problem is to design a source which is small, light, efficient and spectrally indistinguishable from the bomber.

To obtain the desired output, a small source must be operated at high temperature, and vice versa. But a high temperature source will exhibit a markedly different spectral distribution than a low temperature source. This difference is, in principle, sufficient to allow discrimination between a simulated IR radiation and the radiation from the bomber.

The most convenient source of energy for the simulated IR source is jet engine fuel. The flame may not be visible, but can be used to heat a high emissivity radiating surface such as oxidized stainless steel with an emissivity of 0.9. This will insure that much of the radiation will be greybody emission instead of band emission. The hot exhaust gases will give some band emission.

The size of the radiating area will depend on the desired output and oper-
ating temperature. Measurements of the power radiated by large bombers
indicate about 600 watts per steradian in the 2 to 2.7-micron band, while
the B-47 emits about 200 watts per steradian under the same conditions
(Reference 19). These results were obtained at cruising speeds with tail
pipe temperatures in the neighborhood of 400° C.

The fuel requirements depend strongly upon the efficiency with which
thermal energy can be converted into radiant energy. Assuming an efficiency
of 10 percent and that the total radiated power is π times the power radiated
per steradian along the tail axis, the required weight of fuel per hour to
simulate a B-52 is 30 lb/hr for the 2 to 2.7-micron band and 280 lb/hr for
the 2 to 5-micron band. It is also assumed that 15 percent of the received
radiation is effective in producing a signal in the detector and that a hydro-
carbon fuel liberating about 15,000 BTU/lb is used.

20.8 Conclusion

To conclude this chapter on decoy simulation, we shall summarize the
relationship between identification methods and simulation techniques, and
compare the decoy with another device for diluting the defense.

It is assumed that the decoy carries simulation equipment which is ade-
quate against the simplest technique of identification; the break-in range.
The various methods for identifying the decoys are summarized in Table 20-IV.
The top row is a list of different methods for identifying decoy targets. The
left column is a list of techniques which can be used to counter the identi-
fication methods. The pairing of an identification method and a counter
technique is indicated. For example, a possible counter technique is an
examination of the fluctuations of target echoes. This method can be coun-
tered by applying noise modulation to the repeater. In this manner, for each
identification method, a counter technique exists to null the effectiveness of
the method.

Decoy aircraft, employed to saturate the defenses, are often compared with
false target generators. These devices place false targets on defense radars by
purely electronic means. Although both techniques accomplish the same result
they differ in the confidence ascribed to the countermeasure. In the case of
false target generators, there are simple modifications of existing radars
which blank out the false targets. The false target generators cannot be
modified to eliminate this difficulty. They are therefore considered as a
lower confidence countermeasure than the decoys. In the case of decoys, the
methods for identification are more complex, and furthermore, they can be
nulled by modification of the decoy. Consequently the decoys are relatively
high confidence countermeasures.

TABLE 20-IV. SUMMARY OF SIMULATION TECHNIQUES

Counter Techniques	Identification Methods						
	1 Detection of Time Sharing Repeaters	2 Response to Circular Polarization	3 Detection of Jammer Radiation	4 Average Echo Amplitude	5 Fluctuation of Target Echo	6 IR Energy and Spectral Distribution	7 Frequency Dependence of Target Echo
Time Sharing Repeater on Bomber	X						
Small Jammer on Decoy			X				
Circularly Polarized Corner Reflectors		X	X				
Active Passive Simulation of Broadside Echo				X			
IR Simulator						X	
Noise Modulations of Repeaters					X		X
Interaction Between Corner Reflector					X		X

REFERENCES

1. Hult, J. L. and W. B. Graham, "Some Promising Countermeasures Against Enemy Air Defense Systems," The RAND Corporation, RM-1144, September 1953. (SECRET)

2. "MX-2223 Simulation Studies," Parts I, II, III, The Johns Hopkins University Radiation Laboratory Tech. Report AF-15, Contract No. AF 33(616)-68, March 1955, September 1955, Contract No. AF 33(616)-3374, October 1956. (SECRET)

3. "Technical Capabilities of Soviet Air Defense Weapons," Air Technical Intelligence Center, TIS-ES-56-3, Proj. No. 17002, 15 April 1956. (SECRET)

4. "Handbook of Soviet and Satellite Radar Equipment," Air Technical Intelligence Center, TIH-EL-1, Proj. No. 20059, May 1956. (SECRET)

5. "Analysis of the YO-YO and Associated Electronic Equipment," Air Technical Intelligence Center, TIS-EL-56-1, Proj. No. 23002, 25 September 1956. (SECRET)

6. "Moscow Surface to Air Guided Missile Systems," Air Technical Intelligence Center Study TIS-GM-56-1. (SECRET)

7. Kennaugh, E. M., "A Corner Reflector for Use With Circularly Polarized Radars," Ohio State University Antenna Laboratory Report 601-8, Proj. 5-(4-460), January 30, 1956.

8. "Echo Measurements of the B-47 Aircraft, "OSURF Data Set No. 3, Ohio State University Antenna Laboratory, June 12, 1952. (CONFIDENTIAL)

9. Schensted, C. E., J. W. Crispin and K. M. Siegel, "Radar Cross Sections of B-47 and B-52 Aircraft," Willow Run Research Center Report No. 2260-1-T, AF 33(616)-2531, The University of Michigan, August 1954. (SECRET)

10. Van Vleck, Bloch, and Hammermesh, Theory of Radar Reflections from Wires or Thin Metallic Strips, *J. Appl. Phys.*, March 1947.

11. Wilcox, C. H., "The Refraction of Plane Electromagnetic Waves by a Luneburg Lens," Lockheed Aircraft Corporation Report MSD 1802, Missile Systems Division, Van Nuys, California, June 1956.

12. Technical Bulletin 6-2-3A, Ecco Reflector, Emerson Cuming, Inc., 869 Washington St., Canton, Massachusetts.

13. Spenser, R. C., "Optical Theory of the Corner Reflector," Massachusetts Institute of Technology Radiation Laboratory Report No. 433, March 1944.

14. Bonkowski, R. R., C. R. Lubitz and C. E. Schensted, "Cross-Sections of Corner Reflectors and Other Multiple Scatters at Microwave Frequencies," Willow Run Research Center UMM-106, The University of Michigan, October 1953.

15. Cohen, M. H. and R. C. Fisher, "A Dual-Standard for Radar Echo Measurements," *IRE Trans. on Ant. and Prop.*, Vol. AP-3, July 1955.

16. Chu, L. J., "Analysis of Window and Related Matters," RRL-4, Harvard University, October 1942.

17. Marcum, J. I., "A Statistical Theory of Target Detection by Pulsed Radar," The RAND Corporation, RM 754, 753. (SECRET)

18. Swerling, P., "Probability of Detection for Fluctuating Targets," The RAND Corporation, RM 1217, March 17, 1954. (SECRET)

19. The Johns Hopkins University Radiation Laboratory, "Symposium on Military Applications of Infrared Physics," Tech. Report AF-31, p. 89. (SECRET)

21

Characteristics of Infrared Radiation

M. D. EARLE

21.1 Introduction

Infrared refers to the portion of the electromagnetic spectrum lying between the visible and the short wavelength microwaves. Sir William Herschel discovered the presence of energy in this region of the spectrum in 1800 while exploring the distribution of energy in the solar spectrum. Using a mercury-in-glass thermometer as a detector, he found that the thermometer did not reach its maximum reading until it had been moved a considerable distance beyond the red end of the visible spectrum.
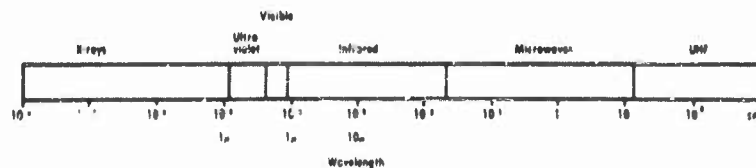
FIGURE 21-1 Relation of infrared to other regions of the electromagnetic spectrum

Figure 21-1 shows the position of infrared in the electromagnetic spectrum. The usual unit for the measurement of infrared wavelength is the micron (μ), which is equal to 10^{-4} centimeters. The infrared spectrum can be roughly divided into three parts; the near infrared from .8 to 1.2 microns, the intermediate infrared from 1.2 to 7 microns, and the far infrared beyond 7 microns. The region from 1.2 to 7 microns is the most important from the

military standpoint as this is the region in which the photoconductive detectors are useful. This is also the region into which most of the infrared radiation from military targets falls.

Since all objects at a temperature above absolute zero emit infrared radiation to some extent, many objects of military importance are unavoidably good infrared targets. The intensity of the radiation increases and the position of the peak intensity moves toward shorter wavelengths as the temperature is increased. If one imagines a material such that all the radiant energy falling on it is absorbed, regardless of its wavelength, it would be found that the power radiated by this body per unit area of surface and the spectral distribution of this radiated power is a function only of its temperature. Such an object is called a blackbody. While ideal blackbodies do not exist, a very close approximation to one is a uniformly heated enclosure with a small aperture through which the radiation emanates.

The spectral distribution of energy in blackbody radiation is expressed by the following formula, known as Planck's radiation law:

$$W_\lambda \, d\lambda = \frac{2\pi c^2 h}{\lambda^5 (e^{ch/\lambda kt} - 1)} \, d\lambda \tag{21-1}$$

where: $W_\lambda \, d\lambda$ is the radiant emittance within the wavelength band $d\lambda$.

       c is the velocity of light
       h is Planck's constant
       λ is the wavelength
       T is the absolute temperature
       k is Boltzmann's constant

If the above expression is divided by $d\lambda$, the spectral radiant emittance W_λ is obtained. This is the radiant emittance per unit wavelength interval. (It can be expressed in watts per square centimeter per micron, for example.)

If Eq (21-1) is differentiated with respect to λ and placed equal to zero, the value of the wavelength corresponding to the maximum value of the spectral radiant emittance is obtained. The resulting expression is

$$\lambda_{max} = \text{constant}/T \tag{21-2}$$

This is known as Wien's displacement law. It shows that the maximum of the blackbody spectrum shifts toward shorter wavelengths as the temperature is increased.

If $W_\lambda \, d\lambda$ is integrated over all values of λ the total radiant emittance is obtained. This is known as the Stefan-Boltzmann law.

$$\int_0^\infty W_\lambda \, d\lambda = 2\pi \, c^2 h \int_0^\infty \frac{d\lambda}{\lambda^5(e^{ch/\lambda kt} - 1)} \, d\lambda = \text{constant}(T^4) \quad (21\text{-}3)$$
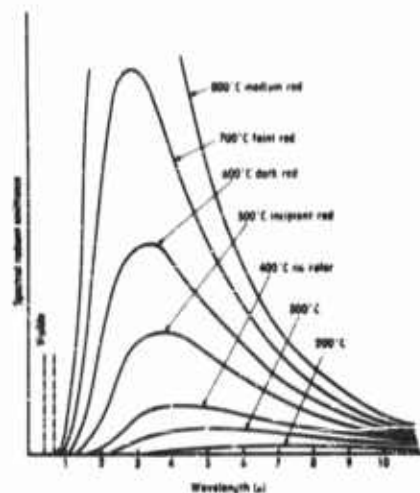
FIGURE 21-2 Relative spectral radiant emittance of blackbodies at several temperatures

In Figure 21-3 the relative spectral radiant emittance is plotted as a function of wavelength for blackbodies at several different temperatures. This figure also shows the shift of the position of the maxima of these curves toward shorter wavelengths as the temperature is increased.

A graybody is defined as one which when heated has a radiation spectrum the same shape as that of a blackbody, but whose spectral radiant emittance is reduced by a constant factor throughout the spectrum. The emissivity of such a body is defined as the ratio of its radiant emittance to that of a blackbody at the same temperature.

Many gases when heated are also emitters of infrared radiation. If the vibration and rotational energy of a gas have the proper values, the gas will absorb infrared radiation in certain regions of the infrared spectrum. Since a heated gas emits radiation of the same frequencies that it absorbs, these gases will also radiate in the infrared region of the spectrum. For this reason, this type of radiation is not continuous as is that from a black or a graybody but is composed of lines. These lines are grouped into bands in various spectral regions.

21.2 Scattering and Absorption by the Atmosphere

In any contemplated use of infrared radiation, the attenuating characteristics of the intervening media must be taken into consideration. For work at low altitudes the carbon dioxide and water vapor in the atmosphere are the most important absorbing constituents. Figures 21-3 and 21-4 show the fractional transmission of infrared as a function of wavelength over a path of 1 nautical mile near sea level, containing 17 mm of precipitable water vapor. The amount of precipitable water vapor is defined as the thickness
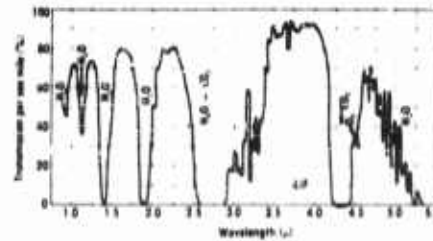
FIGURE 21-3 Atmospheric transmission in the
1 to 5.5 micron region (Ref. 1)
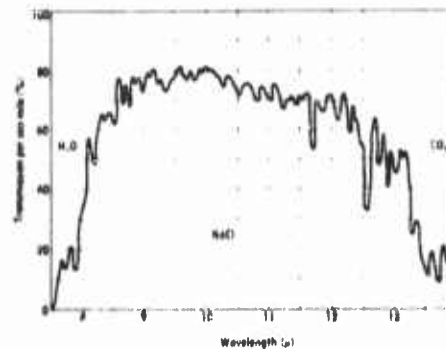
FIGURE 21-4 Atmospheric transmission in the
7.5 to 14 micron region (Ref. 1)

(pathlength) of water which would be present if all the water vapor in the path were condensed into liquid water.

Figures 21-3 and 21-4 show several regions of high transmission called "windows". The absorption bands which separate these windows are due either to CO_2 or H_2O, as indicated. The absorption band centered near 2.7 microns is due to CO_2 and H_2O.

The concentration of CO_2 in the atmosphere is about .033 percent (Reference 2) by volume. This value remains almost constant with altitude. The absolute amount of CO_2 in a path of given length would therefore be proportional to the density.

For radiation in the spectral regions from 1.9 to 2.1 microns, 2.2 to 2.6 microns, and 2.9 to 4.1 microns there is very little absorption by CO_2. There are strong absorption bands at 2.7 and 4.3 microns. The attenuation over a 5-kilometer path (3.1 miles, a reasonable range for a small air-to-air missile) due to CO_2 in the atmosphere for 2.7 micron radiation would vary from 94 percent at sea level to about 43 percent at 40,000 feet. For radiation of 4.3

microns there is almost complete absorption even at 40,000 feet for a path-length of 5 kilometers. Due to the broad transmission regions mentioned above, the absorption by CO_2 in the 1.2 to 7 micron region is not severe, even at sea level, and becomes even smaller as the altitude increases.

The variation of H_2O vapor with altitude is shown in Table 21-I (Reference 3). These are average values for the eastern part of the United States.

TABLE 21-I. Variation of Water Vapor with Altitude

Altitude (ft)	Absolute Humidity (g/m^3)
5000	14.0
10000	8.8
15000	4.4
20000	2.3
30000	1.2
40000	0.01

Computations using the data shown in Table 21-I indicate that, for the spectral regions from 1.9 to 2.7 microns, 2.7 to 4.3 microns, and 4.3 to 5.9 microns (Reference 4), the attenuation due to water vapor alone over a 5-kilometer path at sea level would be approximately 50, 50, and 88 percent respectively. The attenuation at 40,000 feet for the 1.9 to 2.7 micron and 2.7 to 4.3 micron region would be negligible whereas for the 4.3 to 5.9 micron region it would be only about 2 percent for a 5-kilometer path.

While CO_2 and H_2O are the principal absorbers in the atmosphere, other constituents, such as nitrous oxide, N_2O; methane, CH_4; carbon monoxide, CO; and hydrogen deuterium oxide, HDO (Reference 5) have absorption bands in the infrared region but play minor roles in military applications. Ozone has a relatively strong absorption band at 9.6 microns and weaker ones at about 4.7 and 14.1 microns. Ozone reaches its maximum concentration at about 80,000 feet (see Figure 22-12) and the effect of its presence in the infrared spectral region is small. This is particularly so in the region of sensitivity of the commonly used photoconductors.

Some attenuation of infrared radiation is caused by scattering due to the presence of haze in the atmosphere. This effect is most pronounced at low altitudes and at the shorter infrared wavelengths. Fog and clouds attenuate infrared radiation very strongly. It is estimated (Reference 3) that the transmittance of a typical cloud in the near infrared would be only one percent for a pathlength of approximately 400 feet. For this reason, the use of infrared devices at low altitudes is very uncertain, whereas at altitudes above about 30,000 feet, weather and water vapor are less important factors.

21.3 Comparison of Infrared Devices With Those Employing Microwave and Visible Radiation

Some of the advantages of infrared over microwaves are: 1) Due to the short wavelengths involved, much greater resolution can be obtained with infrared devices which use very much smaller "aper res" than with microwave devices. 2) Since many infrared systems are passive, detection of their presence is difficult, increasing the complexity of the countermeasures problem for the enemy. 3) In an active system (e.g. communications) the energy cannot be intercepted by the enemy without getting directly in the beam. 4) The electronic circuits in infrared devices are usually simpler and cheaper than those in microwave devices. Since radiation chopping frequencies used in infrared techniques are in the audio range, complicated high frequency techniques are avoided.

Some of the disadvantages of infrared as compared to microwaves are: 1) Clouds and water vapor greatly reduce the effectiveness of infrared systems. 2) Infrared sources and detectors cannot be "tuned" as sharply as microwave devices. This means that much power is wasted because it is of a frequency to which the detector is insensitive. 3) Passive infrared systems do not give range information. 4) Background radiation, particularly in daylight, is often troublesome.

It is difficult to compare infrared and microwave systems with respect to maximum range since this depends on many factors. In general, the area of a target for radar reflection is larger than the area for infrared radiation (i.e., the radar cross section of an aircraft compared to the area of its tail pipes). However, neglecting absorption and scattering, the signal is proportional to the inverse fourth power of the range in reflecting (active) systems, whereas it drops off as the inverse square of the range for passive systems where the target is the source of radiation.

Infrared has several advantages over visible radiation for military purposes. Some of these are: 1) Many military targets are at temperatures such that considerable infrared is radiated, while often little or no visible radiation is present. 2) Since infrared is invisible to the human eye, greater security is possible. 3) In infrared systems advantage can be taken of spectral filtering to remove much of the background radiation produced by scattered and reflected sunlight, and thus achieve greater daytime capability than would be possible in systems using visible radiation. 4) Infrared is slightly better for penetrating fog than is visible radiation.

21.4 Infrared Detectors

One of the principal contributing factors to the present day effectiveness of military infrared devices has been the great progress which has been made

In the development of high-sensitivity, short-time-constant, photoconductive detectors. The e detectors have their chief usefulness in the intermediate (1.2 to 7 mic ns) infrared region. Photoemissive cells and photographic techniques can be used to about 1.2 microns. Before the advent of the photoconductive detector, exploration of the region beyond 1.2 microns was limited to the use of thermal detectors such as bolometers and thermopiles. These devices were slow and frequently of low sensitivity.

The action of a photoconductor does not depend upon a heating of the detector material as is the case in a thermal detector. It is a quantum effect involving the action of quanta of radiation on the current carriers of the conducting material. For this reason the time constant (defined as the time required for the detector output signal to decay to $1/e$, or approximately 37 percent, of its steady-state value upon interruption of the radiation falling upon it) can be very short. Another fundamental difference between photoconductive and thermal detectors is that the spectral response of photoconductors varies with the wavelength of the incident radiation as compared to the uniform spectral response of "black", i.e., perfectly absorbing thermal detectors.

The chief photoconductive materials presently being used in military infrared hardware are: lead sulfide, PhS; telluride, PbTe; lead selenide, PbSe; and indium antimonide, InSb. Figure 21-5 shows the spectral response of some typical photoconductors. In these curves "Jones' S" (first introduced by Dr. R. Clark Jones) is plotted against wavelengths.

Jones' S is given by the following expression:
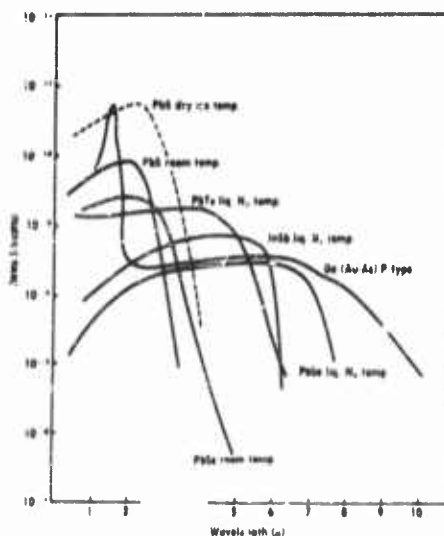
FIGURE 21-5 Spectral response of typical photoconductors (Ref. 3)

$$S = \frac{NEP}{A^{\frac{1}{2}}} (f/\triangle f)^{\frac{1}{2}} \quad (21-4)$$

where: NEP is the noise equivalent power

A is the area (ratio of actual area to unit area to make the equation dimensionally correct)

f is the center frequency (chopping frequency)*

$\triangle f$ is the amplifier bandwidth (should be small compared to f)

S is then the value of the NEP of a similar cell in what is known as "reference condition C." This condition specifies unit cell area and a bandwidth $\triangle f$ between limits f_2 and f_1 such that $f_2/f_1 = e$ (base of natural logarithms). The derivation of Jones' S is based on the assumption that the cell is current noise-limited and that the noise power per unit bandwidth is inversely proportional to the chopping frequency. This condition holds for a number of photoconductive detectors. The chopping frequency should be low enough so that the full cell response can be realized. (The cell sensitivity increases as Jones' S decreases).

It is necessary to cool some photoconductive materials in order that the noise level may be kept sufficiently low. Lead telluride must be cooled to liquid nitrogen temperature (-196°C). Indium antimonide is usually cooled. Lead selenide can be used at room temperature, but its performance is greatly enhanced by cooling; lead sulfide is usually used at room temperature.

Germanium and silicon which have been "doped" with certain impurities are sensitive to wavelengths much longer than the other photoconductors. Detectors made of these materials are now being manufactured and show considerable promise for future use.

Table 21-II shows the typical characteristics of several photoconductors (Reference 3).

Bolometers and thermocouples are still quite useful in the region beyond 6 microns. Great improvements have also been made in these devices in recent years. Bolometers with time constants of less than a millisecond are now available.†

Another type of thermal detector is the Golay cell. This cell consists of a minute cell of gas, which expands upon heating as a result of absorbing radiation. The expanding gas deforms a metallized diaphragm whose displacement is indicated by an optical lever system. The evaporograph is a thermal detector which can be used as an image-forming device in which the degree of evaporation of a film of oil is a function of the quantity of radiation falling upon it.

---

*It is customary to "chop" or interrupt the radiation falling on photoconductive detectors so that the cell signal can be amplified by a-c methods.

†Thermocouples and thermopiles have been used for many years as detectors of infrared radiation. They are relatively slow and frequently insensitive. However, thermocouples of good sensitivity which can be modulated at frequencies of 10 to 20 cycles per second are now available. These are extensively used in spectrometers.

TABLE 21-II. Characteristics of Photoconductors

	PbS 25° C	PbS —78° C	PbTe —196° C	PbSe —196° C	Ge 25° C
Sensitive Area (sq cm)	0.01 to 400	0.01 to 400	1 to 100	1 to 100	0.1 to 100
Dark Resistance (megohms)	0.1 to 10	1 to 100	1 to 100	0.5 to 10	0.0001 to 5
Time Constant (μsec)	5 to 2000	10 to 5000	5 to 250	1 to 150	0.2 to 50
Wavelength of peak response (μ)	1.5 to 2.5	1.9 to 2.7	1.5 to 4.5	2 to 6	1.6 to 1.8
Long Wavelength limit of response (μ) (Response down to 10% of peak phase)	3.2	3.7	6	7	1.9
NEFD** at peak response (10^{-11} watts/cm^2)	5 to 1000	1 to 100	20 to 1000	100 to 1000	10 to 1000

*These data are based on a chopping frequency of 90 cycles per second and a bandwidth of 5 cycles per second.

**NEFD is noise equivalent flux density which is the flux density at the detector which will produce unity signal-to-noise ratio.

21.5 Principles Involved in Infrared Tracking Devices

The principal threat for which infrared countermeasures are needed, is that presented by air-to-air infrared homing missiles to jet bombers. These missiles have been very successful in tests against drones and tow targets carrying flares to provide a suitable source.

The basic optical component of these and most other infrared devices is the reflecting telescope. The use of reflective optics largely eliminates the difficulty, encountered in this region of the spectrum, of finding suitable transmission optics and provides a device free of "chromatic" aberration, i.e., the focal length of the system does not change with the wavelength of the radiation. A typical optical arrangement is shown in Figure 21-6. This resembles the cassegrainian system used in reflecting telescopes. A true cassegrainian telescope has a paraboloidal primary mirror and a hyperboloidal secondary mirror. However, since extremely high image quality is not required in these seekers, some compromise in favor of cheaper components can be made. For this reason, the primary mirror is usually a spherical mirror and the secondary either a plane or a spherical mirror.
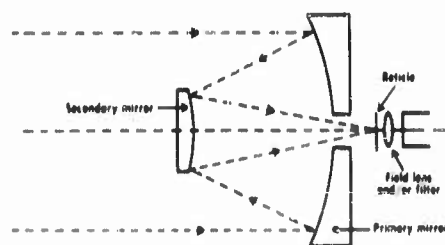
FIGURE 21-6   Optical system of typical infra-
red seeker

The phenomenal success of
this type of missile has been due
to a large extent to two achieve-
ments: the development of
short-time-constant high-sensi-
tivity photoconductive detectors
and the development of means
by which a small target which
occupies only a small part of the
field of view can be distinguished
from its background.

The principal photoconductive detectors have been described earlier in
this chapter. A brief description of background discrimination techniques
will be given here. Since the radiation from the tail pipes of a jet bomber
has its peak in the vicinity of 4 microns and the peak of the sky radiation
occurs at much shorter wavelengths, a considerable improvement in target-
signal to background-signal ratio can be achieved by spectral filtering. For
this purpose an interference filter with its short wavelength cutoff at about
2 microns or a germanium filter whose short wavelength cutoff is at 1.8
microns can be used. The lead sulfide detector has its long wavelength
sensitivity limit at about 2.8 microns. The spectral bandwidth to which the
detector-filter combination is sensitive is about one micron wide and is
located at a value where very little signal is produced by the radiation from
the sky. The target radiation, however, is not excessively attenuated. This
filter can be located as shown in Figure 21-6.

The other technique for target discrimination is based on the difference in
size between the images of the target and the background formed at the
focal plane of the optical system. This method is known as "space filtering".
In one type of seeker, a rotating reticle or "chopper" is placed in the focal
plane of the optical system as shown in Figure 21-6 to interrupt the signal
at a definite frequency so that a-c amplification can be used and better noise
rejection characteristics can be obtained. Suppose the reticle has the form
shown in Figure 21-7. The image of the target is represented by the dot
while the image of the background covers the entire reticle. It can be seen
that the background radiation passing through the reticle to the cell will be
constant regardless of the position of the reticle, the average transmission of
the reticle being about fifty percent. The target radiation will be modulated
while the spoked part of the reticle is passing over the target image but no
modulated signal will be produced while the other half of the reticle is pass-
ing over the target image. In Figure 21-8 the radiation flux reaching the
detector is shown as a function of time (or angular position of the reticle).

FIGURE 21-7 Reticle pattern (Refs. 6 and 7)
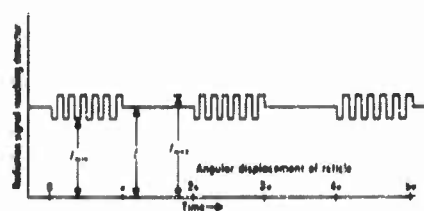
FIGURE 21-8 Radiation pattern as a function
of time for reticle shown in Fig. 21-7 (Ref. 7)

I_{min} represents the flux reaching the detector in the absence of a target (or when the target is completely obscured by an opaque spoke of the chopper). This radiation is due entirely to the background. I_{max} represents the flux reaching the detector with the target present (target image falling on transparent sector of the chopper). I represents the flux reaching the detector when the target image is in the lower half of the reticle. Since the lower half of the reticle has approximately fifty percent transmission for both the target and background radiation, I will be halfway between I_{min} and I_{max}. It can be seen that the difference $I_{max} - I_{min}$ will increase as the target moves outward on the reticle (until spoke width is equal to target image width). This provides a measure of the amplitude of the a-c component of the detector signal while the length of the envelope of each group of pulses will remain constant. When the target image is exactly at the center of the reticle, the target radiation reaching the detector does not change with the angular position of the reticle and the a-c component of the detector signal is zero. After removal of the d-c component, amplification and

rectification of the detector signal, the height of the resulting pulse envelope represents the error signal which makes tracking possible, while the phase relationship with respect to a fixed point on the seeker frame can be used to obtain directional information. A more sophisticated form of reticle is shown in Figure 21-9. In this reticle the error signal increases with radial displacement until the target image reaches the approximate position of the dotted line. After this region is passed, the number of chopping elements decreases with further increase of radial displacement and causes the error signal to decrease also. This technique makes the seeker less susceptible to countermeasures in which a false target is employed since it requires that the false target be near the center of the field of view to have maximum effectiveness. A seeker with this type of response is less likely to be confused in the presence of multiple targets than one with the response previously described. The error signal as a function of error angle (angle between targets and longitudinal axis of the seeker) for a reticle of the type illustrated in Figure 21-7 is shown in Figure 21-10. The error signal curve for the reticle of Figure 21-9 is shown in Figure 21-11.
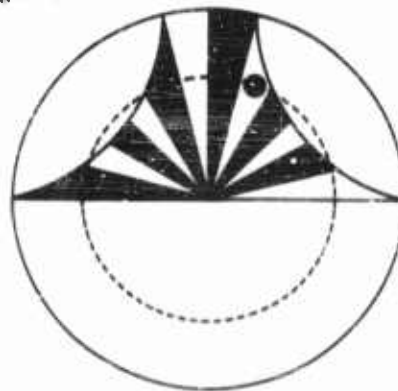
FIGURE 21-9   Improved   form   of   reticle
(Ref. 7)

FIGURE 21-10   Error   signal   versus   error
angle for reticle shown in Fig. 21-7
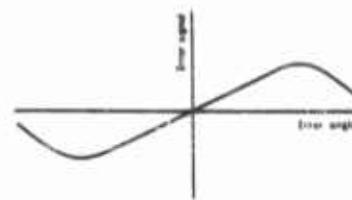
FIGURE 21-11   Error   signal   versus   error
angle for reticle shown in Fig. 21-9

In actual practice a checkered pattern is often used instead of the spoked patterns which have been described. The chief advantage of the checkered pattern over the others is that it shows less response to gradients in the background radiation that tend to lie along the radius of the reticle; a cloud edge, for example.

In one type of infrared seeker the rotating reticle is mounted as a part of the rotating member of a gyroscope. The error signal, through magnetic coupling, produces a torque of the proper phase and magnitude to cause the gyroscope to precess in such a manner that the telescopic system accurately tracks the target. In some systems the primary mirror is also the rotor of the gyroscope.

The first infrared missiles used the lead sulfide cell with a suitable filter and had a useful sensitivity in the region from 1.8 to 2.8 microns. Development of air-to-air missiles employing photoconductive detectors sensitive to longer wavelengths is in progress. This has the following advantages: 1) Many military targets have temperatures such that the peak of their radiation occurs between 3 and 5 microns. 2) The spectral bandwidth of these longer wavelength detectors is wider; hence more power is available. 3) There is less background radiation. 4) These factors result in an increase in the signal-to-noise ratio in spite of the fact that these longer wavelength detectors are usually less sensitive than lead sulfide detectors. Characteristics of some typical air-to-air infrared missiles are listed in Table 21-III.

TABLE 21-III. Characteristics of Some Typical Infrared
Air-to-Air Missiles.

Range	18,000 to 36,000 feet
Speed	Mach 2 to 3
Length	6 to 10 feet
Diameter	5 to 11 inches

21.6 Infrared Characteristics of Jet Aircraft Targets

All aircraft are emitters of infrared radiation. The only ones, however, which will be discussed here are jet aircraft since from a military standpoint they are of much greater importance and since their infrared output, in general, far exceeds that of propeller driven craft.

There are three important sources of infrared radiation in jet aircraft. These are: 1) The radiation from the hot metal parts of the jet engine (turbine and tail pipe). 2) The radiation from the hot gases behind the aircraft in the plume. 3) The radiation from aerodynamically heated surfaces. By far the most important of these is the radiation emitted by the hot metal parts of the engine. The plume radiation is confined chiefly to the spectral

regions in which the principal combustion products, CO_2 and H_2O, have their absorption bands. Radiation due to aerodynamically heated surfaces is negligible for subsonic aircraft, but will probably be a major source of radiation from aircraft such as the Mach 3 bomber. Figures 21-12, 21-13,
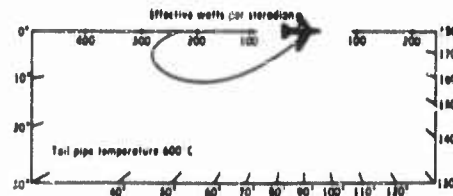
FIGURE 21-12 Radiation pattern for F9F
aircraft (Ref. 6)
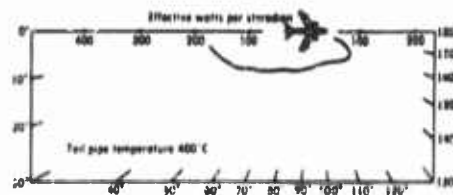
FIGURE 21-13 Radiation pattern for B-47
aircraft (Ref. 6)
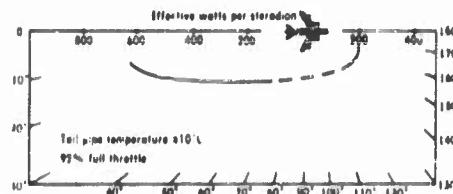
FIGURE 21-14 Radiation pattern for B-52
aircraft (Ref. 6)

and 21-14 show the radiation patterns of an F9F, a B-47, and a B-52 aircraft respectively. The values are shown in effective watts per steradian. This is the power within the 2.0 to 2.6 micron wavelength region which is effective in producing a signal in a seeker equipped with a lead sulfide cell after allowance has been made for atmospheric absorptions, optical filter transmission, and cell response. These patterns were measured at the Naval Ordnance Test Station, China Lake, California. The instrumentation was located on the ground while the aircraft flew in a pattern at low altitude. Dotted portions indicate regions in which measurements were not made.

The pattern for the single jet F9F shown in Figure 21-12 shows a some-what higher maximum value than the B-47. This is due to the much higher tail pipe temperature of the F9F (600°C as compared with 400°C). Some earlier jet aircraft with a more exposed tail pipe showed an almost perfect cosine distribution in their near infrared radiation pattern. The F9F has a less exposed tail pipe and consequently the pattern has been modified some-what (see Section 22.7 on shielding jet engines).

Figure 21-13 shows the radiation pattern for the B-47 (equipped with six J-47 engines) for the same spectral region. The exhaust gas temperature at which this curve was made was 400°C. This temperature would correspond to a throttle setting which would give maximum range if the aircraft were at about 35,000 feet altitude. The speed under these conditions would be about Mach 0.7. At full throttle the output would, of course, be much greater.

Figure 21-14 shows the pattern for a B-52 aircraft equipped with eight J-57 engines. The increased output over the B-47 is due to the following facts: 1) The engines are larger. 2) There are eight of them instead of six. 3) The J-57 engine produces less smoke than the J-47 does. Both these air-craft show considerable forward radiation which may be due to radiation through the engine intakes, or to reflections from the aircraft skin.

The output in this region of the spectrum is due almost entirely to the radiation from the hot metal parts of the engine, very little originating from the plume (Reference 7). The infrared output of a jet engine is strongly dependent on temperature; as has already been shown, the Stefan-Boltzmann law indicates that the total radiated output of a blackbody is proportional to the fourth power of the absolute temperature. For a limited range of wave-length and in the proper temperature region the dependence of radiated output on temperature is much greater than this. This is true because a larger fraction of the total output is in the region of interest, since the maxi-mum of the blackbody shifts toward shorter wavelengths as the temperature is increased. In the 2.0 to 2.6 micron spectral region and for the temperature range 400°C to 700°C, it has been found that the radiated output from a jet engine varies approximately as the seventh power (Reference 7) of the absolute temperature as indicated by an uncooled lead sulfide detector. The following example will illustrate this. Suppose a blackbody to have a tem-perature of 500°C. The total output of this source would be 2 watts per square centimeter. For a temperature of 500°C, 5.5 percent of the total output is in the 2.0 to 2.6 micron region. Thus the output in this band would be 0.11 watts per square centimeter. Suppose the temperature of the black-body to be raised to 600°C. The total output would now be 3.3 watts per square centimeter and of this .27 watts per square centimeter or 8.2 percent would be in the 2.0 to 2.6 micron region. The ratio of those two values is

.27/.11 = 2.45. If the output in this spectral region varied as the seventh power of the absolute temperature, the ratio would be

$$\left(\frac{273 + 600}{273 + 500}\right)^7 = \left(\frac{873}{773}\right)^7 = 2.35.$$

It can be seen that the seventh power law holds reasonably well.

The radiation output values shown in Figures 21-13 and 21-14 for the B-47 and B-52 respectively were taken at throttle settings considerably below full military power. Later measurement (Reference 8) made on the B-47 and B-52 at 35,000 feet under conditions of full military power showed maximum values (in the 1.9 to 2.6 micron region) from the B-47 to be as high as 1000 to 1200 watts per steradian and from the B-52 to be as high as 3000 watts per steradian. These increased values over those which were made at low altitudes can easily be explained by the higher tail pipe temperature involved when the throttles are set for full military power.

The character of the radiation from the hot metal parts of a jet engine resembles that of a blackbody in spectral as well as in spatial distribution.
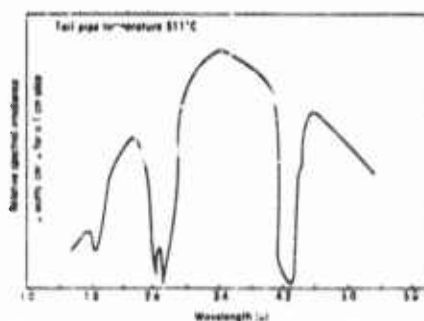
FIGURE 21-15 Tailpipe radiation from the J-17 P3 engine (Ref. 9)

Figure 21-15 shows the relative spectral irradiance of the radiation from the tail pipe of a J-57P3 engine on the ground at close range (120 feet). Except for the strong absorption bands due to H_2O and CO_2 near 2.7 microns and due to CO_2 at 4.3 microns, the spectrum resembles that of a blackbody fairly closely. The peak occurs near 3.4 microns. A blackbody with its peak at this wavelength would have a temperature of 570°C. This is in reasonable agreement with the measured value of 511°C when one considers the difficulty of determining the position of the peak in the measured spectrum due to the absorption of the intervening atmosphere and the combustion products.

Spectral measurements of the plumes of jet engines show that most of the energy is concentrated in the regions near 2.7 and 4.3 microns. This radiation has its origin in the hot CO_2 and H_2O which are the principal combustion products. Since the outer portion of the plume and the atmosphere contain CO_2 and H_2O at a lower temperature than that near the center of the plume,

close to the nozzle, considerable energy at the centers of these bands is absorbed. Plume radiation is not particularly important in the 2.7 micron spectral region, it having been estimated that as much as 95 percent of the radiation in the lead sulfide region (1.8 to 2.8 microns) originates from the tail pipe. The chief significance of the plume radiation is that it has a much broader spatial distribution pattern than does the radiation from the tail pipe. There is considerably more energy from the plume in the 4.3 micron region of the spectrum than in the 2.7 micron region, although it is still much less than that of the tail pipe. The ratio of the output in the 4.3 micron band to that in the 2.7 micron band may vary by almost a factor of 10 and may be as much as 25 or 30. Since the 2.7 micron band radiation falls off more rapidly with temperature than does the radiation at 4.3 microns the ratio mentioned above increases with altitude, because of the lower throttle settings and correspondingly lower temperature usually involved. Total plume radiation drops rapidly with altitude. Measurements have shown that at 40,000 feet the plume radiation in the 2.7 micron region is down to about six percent of its ground level value, while the plume radiation in the 4.3 micron region is down to about 45 percent of its ground level value (Reference 9). The radiation in the 4.3 micron region at 40,000 feet would be about ten percent of the tail pipe radiation. At sea level the plume radiation of the J-57 engine in the region between 4 and 5 microns, for a throttle setting of 93 percent would be about 110 watts per steradian from tail aspect (Reference 3). At 40,000 feet this would drop to about 50 watts per steradian.

The infrared radiation from an aircraft is greatly increased by the use of an afterburner, the output of an engine sometimes increasing by a factor of 50 or more when its afterburner is turned on.

In the case of some supersonic aircraft the radiation from the aerodynamically heated skin will probably be much greater than that from the hot metal parts of the engines and the jet plumes combined. It has been estimated that a six-engine aircraft, flying at a speed of Mach 3 at an altitude of 75,000 feet may have a skin radiation of as much as 60 kilowatts per steradian for the 3.0 to 5.2 micron band in some directions if the skin material is in an oxidized condition (Reference 10).

Afterburner plume radiation is expected to be practically negligible compared to that of the skin. This is due to the fact that at these high speeds there is a great reduction in pressure as the hot gases leave the engine nozzle. This large change in gas pressure is accompanied by a corresponding drop in temperature. The resulting plume temperature behind the aircraft will probably be less than 500 C and the plume radiation in the 4.3 micron region will probably be less than 1 kilowatt per steradian, even when all six engines can be seen by the detector.

The other sources of radiation in this aircraft will be the engine intakes which may have temperatures as high as 300°C and the hot metal parts of the engines. The total output of the six engines will probably not be more than 4 kilowatts per steradian in the 4.3 micron region. The view of the hot metal and gases within the engine will be somewhat restricted by long exhaust nozzles.

21.7 Radiation Characteristics of Air-to-Air Rockets

The infrared radiation characteristics of rockets are also of interest in a discussion of infrared countermeasures. This is because use can be made of the radiation from the rocket motors of missiles in warning systems. (See section on engagement warning systems, Section 22-10.) Ultraviolet measurements on rockets have also been made in this connection.

Table 21-IV shows radiation characteristics of air-to-air rockets in the infrared in kilowatts per steradian at nose-on aspect (Reference 11). The number in parentheses following the average values indicate the burning time in seconds over which the average was taken. Those following the maximum values indicate the time at which the maximum occurred. These values are from the nose-on aspect. Measurements were made at a range of 500 feet. No corrections have been made for atmospheric absorption. These measurements were made at the Naval Ordnance Test Station, China Lake, California, where the water content of the atmosphere is quite low.

TABLE IV.   Radiation Characteristics, Air-to-Air Rockets.

Type Rocket	1.8–2.8 μ		1.8–7 μ	
	average	max	average	max
FFAR MK 2 Mod 0	0.10 (1.6)	0.16 (0.7)	0.30 (1.9)	0.50 (3 0)
FFAR Experimental	1.2 (0.7)	2.2 (0.5)	1.0 (1.5)	2.0 (0.4)
FFAR 107 B	0.15 (1.6)	0.47 (0.3)	0.35 (1.5)	1.1 (0.9)
HVAR	1.0 (1.1)	1.7 (0.4)	3.2 (1.5)	6.0 (0.5)
Zuni	6.8 (1.2)	0.0 (0.3)	19.0 (1.6)	34.0 (0.9)
HPAG	0.05 (1.5)	0.30 (0.1)	0.16 (2.2)	0.28 (0.2)

TABLE 21-V.   Radiation Characteristics of Several Rockets in the Ultraviolet

Type Rocket	Motor	Minimum	Maximum
FFAR	MK 1 Mod 3	0.1	0.13
FFAR	MK 2 Mod 0	3.0	8.0
FFAR	Exp X-12	12.0	15.0
HVAR	MK 10 Mod 6	18.0	27.0
HPAG	123 G	0.2	0.8

Table 21-V shows the radiant intensity of the 2600 Å to 2900 Å ultraviolet region of the spectrum for several rockets (Reference 12). These measurements were also made at nose-on aspect at a range of 500 feet. Values are in milliwatts per steradian.

21.8 Radiation Characteristics of Ground Targets

Ground targets and ground installations are also emitters of infrared radiation. However, there is not a great deal of quantitative data available on their radiation characteristics. Ground targets, being, in general, at lower temperature than aircraft targets, radiate at correspondingly longer wavelengths. Since the radiation from ground targets must pass through the atmospheric layer near the earth (which contains considerable CO_2 gas and H_2O vapor), the atmosphere windows shown in Figure 21-3 and 21-4 are important considerations. Much of the ground radiation passes through the 4.5 to 5 and the 8 to 13 micron windows. For this reason, the photoconductors sensitive to the longer wavelength (such as PbSe and PbTe), and thermal detectors such as bolometers are much more useful than lead sulfide cells. It is conceivable that the radiation from ground targets such as cities and military installations might be used for terminal guidance of air-to-ground or surface-to-surface missiles. So far the effort in this direction has been chiefly limited to short range tactical applications such as infrared guidance for anti-tank weapons or guided bombs.

It has been estimated that a tank which would probably have a temperature in the range of 100 C should be detectable by infrared means at ranges of 1 to 2 miles (Reference 3) during the day and 2 to 4 miles at night.

Another air-to-surface application of infrared is the possibility of detecting submarine wakes, due to the fact that there is a slight temperature gradient between the wake and the surrounding water.

Since the temperature and emissivities of various types of terrain have different values, it has been possible using infrared techniques, to obtain very accurate high resolution maps of ground installations. Concrete runways, for instance, are easily distinguishable from the ground and grass adjacent to them. Vehicles with their engines running   other objects at different temperatures from the background such as antenna, have been clearly detected. It should be mentioned, however, that any device operating by means of infrared radiation from ground targets is rendered useless by heavy rain or cloud cover.

REFERENCES

1. Gebbie, H. A., *et al*, Atmospheric Transmission in the 1 to 14 Micron Region, *Proc. Roy. Soc.*, Vol. A 206, pp. 87-107, 1951.

2. Kupler, G. P., "The Earth as a Planet," University of Chicago Press, 1954.

3. Ballard, S. S., L. Larmore, and S. Passman, "Fundamentals of Infrared for Military Applications," RAND Report R-297, March 31, 1956. (CONFIDENTIAL)

4. Elder, T. and J. Strong, The Infrared Transmission of Atmospheric Windows, *J. Franklin Inst.*, Vol. 255, No. 3, pp. 189-208, March 1953.

5. Howard, J. N., Atmospheric Transmission in the 3.5 Micron Region, *Proc. IRIS*, Vol. 2, No. 1, pp. 59-75, June 1957.

6. Radiation Laboratory, Report on Symposium on Military Applications of Infrared Physics, January 25 and 26, 1956, Johns Hopkins University, Tech. Report AF-31, June 1956. (SECRET)

7. Bloermin, L. M., "A Review of Electro-Optical Countermeasures at NOTS," NAVORD Report 4984, November 30, 1955. (SECRET)

8. Haller, Raymond and Brown, Inc., Proceedings of the Symposium on Optical Radiation from Military Airborne Targets, Report No. AFCRC TR 58-146, Contract AF 19(604)-2541, April 30, 1958. (SECRET)

9. Eastman Kodak Company, Jet Engine Exhaust Survey Report, EK/NOD, EM 937, Contract NORD 12157, February 3, 1956. (CONFIDENTIAL)

10. Gelinas, R. W., "An Estimate of the Infrared Radiation from a Supersonic Bomber," RAND Report S-75, March 5, 1958. (SECRET)

11. Radiation Laboratory, Quarterly Progress Report on Infrared Countermeasures, Part II, AF 3-7, Contract AF 33(616)-3374, Johns Hopkins University, October 1, 1957. (SECRET)

12. Eastman Kodak Company, Quarterly Report EK-D-97-Q6, Contract AF 33(600)-32197, and AF 33(615)-513e, March 10, 1958. (SECRET)

22

Infrared Countermeasures Techniques

M. D. EARLE

22.1 Introduction

Infrared countermeasures for aircraft are necessary to degrade the effectiveness of enemy infrared guided missiles and reconnaissance devices. This is particularly true of strategic bombing aircraft which radiate large quantities of infrared and which operate at altitudes where weather is seldom an interfering factor.

There is a fundamental difference between infrared and radar countermeasures in that infrared devices cannot be jammed in the usual sense. Infrared homing weapons are passive, that is, they do not depend on reflected radiation from an active source for their operation. They operate solely on the radiation which originates from their intended target. For this reason almost all infrared countermeasures are based on two basic concepts. These are: the false target and the suppression of radiation reaching the detector of the seeker or reconnaissance device. The false target may have a number of different forms. It may be a strong source ejected from the aircraft such as a flare, it may consist of modulated beacons on the wingtips or a source which is towed behind the aircraft.

Techniques which are based on the principle of radiation reduction include the shielding and cooling of jet engines and the use of smokes and screening agents. In the case of airborne targets such as a bomber type aircraft the evasive maneuver may be included in the category of radiation reduction. Since it is not in general possible for a heavy aircraft to change its course

rapidly enough to outmaneuver a short range air-to-air missile, the chief purpose of evasive tactics is to present a less favorable aspect to the approaching missile. The various countermeasure techniques are discussed more fully below.

22.2 False Targets Launched From Aircraft

The false target has been shown to be an effective countermeasure for the protection of subsonic jet aircraft against infrared homing air-to-air missiles. The false target in its common form consists of a pyrotechnic flare similar to those which have been used for illumination and identification purposes. These flares consist of a fuel such as finely powdered magnesium or aluminum and an oxidant. The fuel and oxidant must be uniformly mixed and pressed in a mold to a density which allows the flare to burn uniformly during its life. One such flare is composed of magnesium as the fuel and sodium nitrate as the oxidant. A stoichiometric flare made of these materials would be about 42% magnesium and 58% sodium nitrate by weight. The chemical reaction involved is:

$$5 \, Mg + 2 \, NaNO_3 \rightarrow 5 \, MgO + Na_2O + N_2$$

In practice, the composition usually differs slightly from stoichiometric and a small amount of binder is used, the basic formula for a typical flare (Reference 1) being 47.6% magnesium, 47.6% sodium nitrate and 4.8% binder. Flares of a composition similar to this have been shown to have an output of 500 watts per steradian (Reference 2) in the 2 to 2.5 micron region (for a ⅓ pound flare with a burning time of 8 seconds).

It is not necessary that the "oxidizing" agent in the flare contain oxygen to maintain the energy releasing chemical reaction. A flare composed of magnesium and Teflon (polytetrafluoro ethylene) has also shown promise as an infrared countermeasures source. Such a flare might consist of approximately equal parts of magnesium and Teflon to yield a peak output of almost 13 kilowatts per steradian in the 1.8 to 5.4 micron region of the spectrum.

The reaction involved in the Teflon flare is represented by the following equation.

$$M + (C_2F_4)_z = MF_n + C_m$$

M is a metal such as magnesium or aluminum, $(C_2F_4)_z$ represents Teflon (polytetrafluoro ethylene). The reaction products are a fluoride of the metal involved and carbon. Atmospheric oxygen will oxidize some of the carbon to CO_2 and possibly combine with the metal constituent to produce

metallic oxides. Detailed data on a flare of this type are shown below (Reference 3):

Composition—Approx. 1:1 ratio of Teflon and magnesium by weight
Shape—Cylindrical
Volume—54.75 cubic inches
Surface area—80.2 square inches
Burning time—22 seconds
Peak power output—12,675 watts per steradian
Output power at 11 seconds—6,337 watts per steradian
Total energy—122,772 watt-seconds
Specific energy output—79.5 watt-seconds per gram
Spectral region—1.8 to 5.4 microns
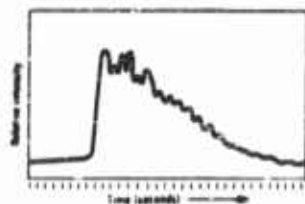
FIGURE 22-1   Burning characteristics of Flora flare (Reference 3)

The output versus time curve for this flare is shown in Figure 22-1.

Another infrared countermeasures source which can be ejected from an aircraft in flight is known as the "Ball of Fire" (Reference 4). This device as originally proposed would consist of a reaction similar to the familiar "thermite" reaction, in which an exchange of oxygen takes place with the release of large quantities of heat. The reactants would be enclosed in a thin graphite shell which constitutes the radiating surface. A large number of possible chemical combinations have been suggested. A few of these are given below:

1. $3Fe_3O_4 + 8 Al \rightarrow 4 Al_2O_3 + 9Fe$
2. $WO_3 + 2Al \rightarrow Al_2O_3 + W$
3. $WO_3 + 3Ca \rightarrow 3CaO + W$
4. $MoO_3 + 3Mg \rightarrow 3MgO + Mo$

A typical device (Reference 5) consists of a 4-inch diameter graphite shell, 0.080 inch thick, which is filled with a mixture of tungsten oxide (WO_3) and aluminum. The reaction involved is that shown in (2) above. It is indicated that such a device when ignited will radiate approximately 1 kilowatt per steradian and will reach a temperature of 2100 K to 2400 K. A typical Ball of Fire is ignited by several hot wire ignition points and has vent holes covering about 10 percent of its area to relieve the internal stress produced by the reaction.

Since Balls of Fire involve a self-contained nonexpanding reaction, it was thought that they should be superior to a burning process in which the

material was dispersed in the form of gaseous reaction products or solid particles in the form of smoke. This would allow the size and weight of the ball of fire to remain approximately constant and thus maintain the weight-to-drag ratio for the countermeasure at a value which would produce better aerodynamic characteristics than could be achieved in a source which was consumed as it burned.

In actual firings of Balls of Fire it has been found that they, like conventional flares, produce a considerable amount of smoke. This smoke consists largely of solid particles of the reaction products produced during burning. This smoke emanates through the vent holes which are necessary to relieve internal stresses. The characteristics of Balls of Fire, therefore, do not seem to differ in any important respects from those of pyrotechnic flares. The flares have the advantage that they are probably somewhat simpler to package and dispense than Balls of Fire would be. Some problems and techniques related to the use of active countermeasure sources are discussed in the next section.

An active countermeasure source should be dominant over the target it is intended to protect. The relative intensity of the decoy source depends on several factors, one of which is the configuration of the engines, i.e., whether it is a single or multiengine aircraft (Reference 6). The ratio of source intensity to target intensity required also depends on the separation rate of the target and source and on the range and direction of approach of the missile. Typical air-to-air missiles use a proportional navigation system and fly what is known as a lead collision course. In this type of guidance the velocity of the target need not be known in order that a collision course be pursued. The diagram in Figure 22-2 shows the principle involved. Suppose that the missile and aircraft are each traveling at a constant velocity, the velocity of the missile, of course, being greater than that of the aircraft. If the missile is on a lead collision course the angle between the line-of-sight of the missile seeker (which points toward the target) and the missile direction will have a constant value. If either the missile or aircraft changes direction or speed the line-of-sight direction will change. An error signal is then produced which enables the seeker to recenter on the target. Signals are transmitted to the control surfaces of the missile so that it is brought back on a collision course. The turning rate of the missile is proportional to the turning rate of the line-of-sight. The value of this ratio is known as the navigation constant.

If the aircraft target launches a countermeasure source such as a flare, the horizontal component of the separation velocity of the flare with respect to the aircraft will be large due to the high drag coefficient of the flare. It can be seen from Figure 22-2 that the missile approaching with the larger
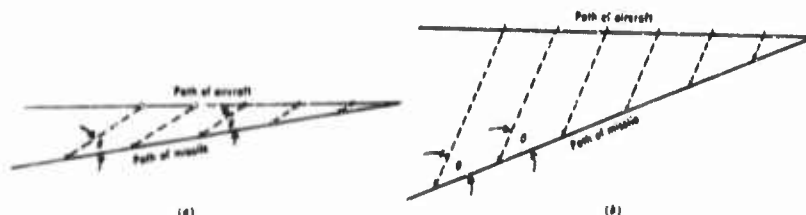
FIGURE 22-2 Lead collision course of air-to-air missile

lead angle will have a greater tracking rate requirement placed upon it than the missile which is approaching with the smaller lead angle. The separation velocity of the flare might be so great that for the condition shown in Figure 22-2b, the missile could not follow it. In this case the flare would move out of the field of view while the target could still be seen by the seeker, allowing it to recenter on the target. This is an unlikely situation even for missiles launched at large angles off the tail since presently operational air-to-air missiles have maximum tracking rates of about 8 to 10 degrees per second. As long as the speed and direction of both missiles and target do not change, the missile seeker need only track at a low rate after the missile is on a lead collision course. This suggests the possibility of constructing missile seekers with low maximum tracking rates as a counter-countermeasure against flares which have a large component of velocity perpendicular to the line-of-sight to the missile However the maximum tracking rate would have to be sufficiently high to allow the missile to follow any evasive maneuvers which the target aircraft might attempt.

If the missile is locked on the target and is to follow a launched source, the missile is required to track actively. As the separation rate of the source increases, as seen by the missile, somewhat greater tracking signals are required, which means that more intense sources are needed. The relationship between relative source intensity and separation rate is shown in Figure 22-3.

These difficulties indicate that the separation rate of target and decoy source should be small. On the other hand, however, the separation rate must be great enough to provide a safe miss distance between flare and target as the missile approaches the flare. If some form of radiation shielding is used, which narrows the radiation pattern of the target and forces the enemy to attack at small lead angles, the countermeasure problem is simplified considerably since such carefully controlled separation rates are no longer required. The shielding problem is discussed more fully later in this chapter.

Another factor of importance in the dispensing of an infrared countermeasure source is the build-up time of the source to full intensity. It is important
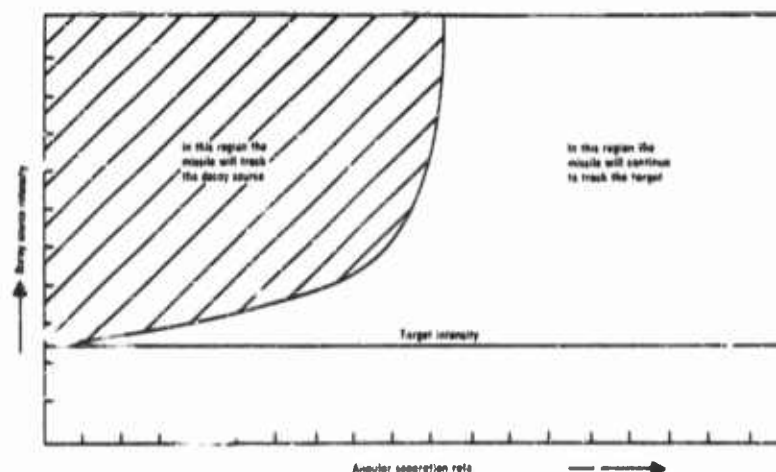
FIGURE 22-3   Relationship between decoy and target intensity (Reference 3)

that the source reach full intensity while it and the target are still unresolved by the seeker system. This is due to the fact that seeker systems are often designed so that a source far from the center of the field of view is less effective than a source of the same intensity nearer the center (see Figure 21-11). Another reason for the source to reach full intensity while near the center of the field of view is that a source near the edge of the field may move completely out of the field of view before the seeker can shift over to it.

In some of the early tests of pyrotechnic flares, ignition took place after the flare had been ejected from the launching tube. It was found, however, that at high altitudes flares frequently did not ignite. This is probably due to the cooling caused by the rapid expansion of the combustion products. This problem was solved by igniting the flare while it was still confined in the launching tube and allowing it to be expelled by the gases produced by the chemical reaction. This also guarantees that the flare reaches full intensity while in the immediate vicinity of the target. The flare is launched within a few milliseconds after the squib has been activated.

Two other important factors to be considered in the tactical employment of infrared countermeasures sources are the burning time and launching interval. If a sufficiently accurate and reliable missile launch detector or engagement warning system were available, the countermeasure source should be launched as soon as possible after the enemy missile has been launched. This would give the countermeasure source the maximum amount of time to act on the enemy missile after the source could be resolved as separate from the launching aircraft. The burning time of the flare should be at least as long

as the expected flight time of the missile. This would insure that the missile could not lock back into the aircraft after the flare had burned out.

In the absence of a missile launch detector, if it is known that an enemy interceptor aircraft is in a position from which infrared missiles might be launched, the time interval between successive flare launchings should not be greater than $t_F - t_m$, where t_F is the flare burning time and t_m is the time required for the flare to reach a distance from the aircraft which would represent a safe missile miss distance. This means that one flare always reaches a safe missile distance before the preceding flare burns out. The safe missile miss distance depends on the fuze and warhead. It should be at least 20 or 30 feet from the nearest part of the aircraft for a contact fuzed missile and greater for a missile equipped with a proximity fuze.

If the quantity $t_F - t_m$ previously mentioned is much greater than either the missile flight time or the expected interval between missile launchings, one of these latter two quantities, whichever is less, should be the controlling factor in determining the countermeasure source launching interval. The minimum interval between missile launchings would probably be determined by the burning time of the rocket motor (approximately 1 to 1.5 seconds for U. S. air-to-air rockets) since the launchings of a missile before the rocket motor of the previous one has burned out would allow the latter missile seeker to track the missile preceding it.

A countermeasure source burning time of 8 to 12 seconds at an intensity which is dominant over the aircraft seems adequate. If the average separation velocity of such a flare during the first second is of the order of 100 feet per second, a launching interval of from 5 to 10 seconds should be satisfactory.

22.3 Flare Dispensing Equipment

Experimental flare dispensers (Reference 7), suitable for use on subsonic aircraft have been built. The basic unit is a flare battery of (Figure 22-4) dimensions $5\frac{1}{2} \times 12 \times 12\frac{1}{2}$ inches consisting of 17 tubes or flare containers. Each tube can contain from one to five flares depending on their lengths. Firing squibs, as shown in Figure 22-5, are placed along the length of each tube and are used for firing the flares in sequence. The flare battery is expendable and fits into a box which is permanently mounted flush with the surface of the fuselage. Electrical connections are made automatically when the battery is plugged into the box. An intervalometer within the aircraft is provided so that the total number of flares and the time interval between their firing can be controlled. The tubes are made sufficiently strong to prevent any damage if one of the inner flares should accidently ignite before the outer ones had been expelled.

FIGURE 22-4   Flare dispenser battery

FIGURE 22-5   Flare containers showing firing squib connections

22.4 Problems Involved When Devices Sensitive to Longer Wavelengths are Used

The previous discussion assumed missiles with seekers sensitive in the 2 to 2.8 micron (PbS) band. The following computations illustrate the increased requirements which will be placed upon countermeasure sources when longer wavelengths are used.

Suppose the target is a B-47 whose tail pipe radiation has the spectral distribution of that from a graybody at 425°C with a total output over all wavelengths of 4200 watts per steradian. A graybody (or blackbody) at these temperatures will have 5.4 percent of its output in the 2 to 2.8 micron band. The intensity of this source in the 2 to 2.8 micron band is therefore 227 watts per steradian.

Now suppose a countermeasure source has a total output over all wavelengths of 2000 watts per steradian and the spectral distribution of a graybody at 2300°C. A graybody at this temperature has 16.7 percent of its output in the 2 to 2.8 micron spectral region. Its output in this wavelength band is therefore 334 watts per steradian. If the seeker system were sensitive only in the 2 to 2.8 micron band, this source might be an adequate countermeasure.

Now suppose the same target was being tracked by a missile whose seeker had its sensitivity in the 3 to 6 micron band. The 425°C graybody target would have 43.8 percent of its output in this region. The intensity in the 3 to 6 micron region would therefore be 1840 watts per steradian. The 2000 watts per steradian, 2300 C countermeasure source, would have 12.7 percent of its output in the 3 to 6 micron region. Its intensity in this spectral region would then be 254 watts per steradian. Thus this source would be entirely inadequate when used against a missile whose seeker is sensitive to the 3 to 6

micron band. The situation might be slightly worse than outlined above since there would also be some 4.3 micron CO_2 plume radiation to which the longer wavelength seeker would respond.

22.5 Low Temperature Decoy

The brief analysis of Section 22.4 points out the need for a suitable low temperature decoy (Reference 8), i.e., one whose temperature is close to that of the target. This type of decoy has an inherent disadvantage, however. Since its temperature is much lower than the sources previously discussed, its radiating area must be much greater in order that its output be sufficiently large. It has been proposed that such decoys could be launched from a bomber or decoy bomber and be supported from a small parachute. The system might also be equipped with a collapsible corner reflector which would unfold on launching and allow the devices to act on a radar countermeasure as well. One of the chief problems appears to be the difficulty of carrying enough such sources on board an aircraft. The sources have to be quite large to provide enough radiating area and to allow a sufficiently long burning time. Since these sources would remain suspended for a considerable period of time they should have a correspondingly longer burning time than flares or Balls of Fire.

One of the chief reasons why a low temperature decoy may become very important is that it is possible to build seekers which reject sources whose spectral distribution is greatly different from that of the target.

The two-color chopper or reticle is an example of such a device. Such a chopper might consist of a wheel or disc composed of alternate sectors of two different materials of characteristics such that predominantly short wavelength radiation, like that scattered from the sky or clouds, would be transmitted equally by each material. Thus, very little modulation of the cell signal would be produced. Now suppose a source such as a jet engine tail pipe with its radiation peak around 2 or 3 microns were viewed. The transmission characteristics of alternate sectors would be quite different for this source and a strongly modulated detector signal would result. Some high temperature sources would thus be very ineffective against a seeker equipped with a two-color chopper.

22.6 Active Sources Attached to the Aircraft

While tests and experience have shown that active sources dispensed from the aircraft are more effective infrared countermeasures than sources permanently attached, brief mention will be made of two other infrared countermeasures techniques involving active sources. These are the blinker counter-

measure and the towed decoy (References 9, 10, and 11). Briefly, the blinker technique involves strong sources placed at the wingtips of an aircraft which can be turned on and off at a suitable rate to produce large disturbances in the path of the missile. In addition to this two dimensional scheme, several three dimensional schemes have been proposed. These are: 1) infrared sources on both wingtips, on the top of the vertical fin and at the lowest extremity of the bomber fuselage, 2) same as case 1) except that the lower source is mounted on an aerial paravane suspended below the bomber, and 3) a single infrared source mounted on a controllable paravane which can be made to revolve about the longitudinal axis of the bomber.

In the two dimensional case mentioned above the maximum possible miss distance might be reasonably large, but since the missile would be caused to cross from one side to the other of the direct path to the bomber, direct hits could still occur. The three dimensional system would cause the missile to fly a spiral trajectory missing the target at all positions.
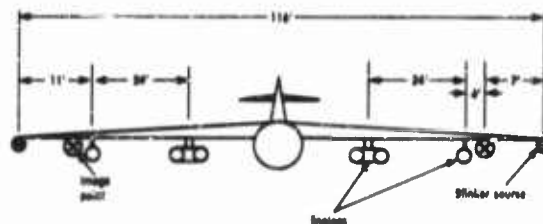
FIGURE 22-6   Blinker system with sources of wingtips of aircraft (B-47) (Reference 12)

Figure 22-6 shows how two wingtip sources might be employed on a B-47 aircraft (Reference 12). Each source would need to be of sufficient intensity to capture the seeker of the missile when the particular source is turned on. Such sources would require a large amount of power. The additional weight and drag of such sources would also be a handicap. A further difficulty is involved in determining the proper modulation or blinker rate. The most effective value for a particular missile will depend on its dynamic characteristics. Since infrared air-to-air missiles have narrow fields of view(approximately $\pm 2°$ to $\pm 3°$), the missile will reach a range where only one blinker will be in its field of view. At this range it will see one blinker go off but will not see the other blinker go on. Since at least one engine will probably still be in its field of view, the missile would probably still home on this engine with only minor perturbations being produced in the path of the missile by the blinker.

The three dimensional blinker system mentioned above would possess

essentially the same disadvantages as the two dimensional system using sources on the wingtips.

The spiraling paravane carrying an infrared source would be a serio aerodynamic handicap to the towing aircraft, would require a large amount of power to operate the infrared source, and would also require a knowledge of the dynamic characteristics of the enemy seeker.

Another active countermeasure technique on which a considerable amount of development work has been done is the towed decoy (Reference 13). This system differs from the spiraling paravane mentioned above in that it does not spiral and is actually a small delta wing towed glider. This glider would probably be towed behind the aircraft at a large distance.

There are a number of disadvantages to such a system; some of the more important are: 1) the aerodynamic handicap and drag; 2) the placing of the decoy at the proper level with respect to the towing aircraft would have to be done very carefully, so that should a missile strike the decoy, the missile or fragments from it would not damage the aircraft; and 3) the towed decoy would protect against only one missile firing if the missile actually hit the decoy.

In view of the difficulties discussed above it does not appear that the blinker countermeasure or the towed decoy are feasible as infrared counter-measures.

22.7 Shielding of Jet Engines

The spatial distribution of the radiation from the aperture of a blackbody follows a cosine distribution. That is, the intensity in any given direction is proportional to the cosine of the angle made between the direction in ques-

FIGURE 22-7   Cosine distribution radiation pattern

tion and the normal to the aperture as shown in Figure 22-7. As has been shown in Figures 21-12, 21-13, and 21-14, the radiation from the tail pipe of a jet engine approximates that of a blackbody in both spatial and spectral distribution. It is conceivable, therefore, that a tail pipe extension or shield could be used with a jet engine which would limit the solid angle from which the hot metal parts could be "seen" by a detector. The shield would only be useful in reducing the effective radiation of the hot metal parts of the engine. Since the plume radiation results from the hot gases aft of the aircraft, the shield does not reduce their radiation outputs. It is esti-mated, however, that only about 5 percent (Reference 14) of the radiation

output of the aircraft in the lead sulfide sensitivity region originates from the plume. Furthermore, the plume radiation decreases rapidly with altitude.

Several requirements must be placed upon such shields if they are to be effective. One of these is that the shield itself must be kept cool both inside and out. The other important requirement is that the interior of the shield must have a low reflectivity in the spectral region of interest. These two requirements are somewhat contradictory since a low reflectivity (i.e. a "black") surface will be a good absorber of radiation and its temperature will rise. If the internal reflectivity of the shield is high it will simply reflect the interior of the blackbody cavity, acting as a "light pipe" and will produce very little change in the radiation pattern.

Some simulation work has been done to evaluate the effectiveness of such shields (Reference 15). The source in these experiments consisted of an acetylene burner into which a metal slug had been placed to simulate the hot turbine blades and interior parts of the jet engine. The combustion products,

FIGURE 22-8   Radiation pattern of jet simulator with no shield (Reference 15)

CO_2 and H_2O approximated those of an actual jet engine. Figure 22-8 shows the radiation pattern of the burner with no shield. It can be seen that the pattern as seen by both PbS and PbTe detectors are close to cosine distributions. The fact that the patterns for PbTe and PbS are so similar indicates that there is little contribution from the gaseous plume. The PbTe detector is sensitive to about 5.5 microns and thus includes the 4.3 micron emission band of CO_2. If there were a large contribution from the plume the two patterns would show greater differences. Using the PbTe detector, the pattern was measured at 400 C, 500 C, and 600 C with very little change in spatial distribution, although, of course, the output increased rapidly with temperature.

Figure 22-9 shows the effect of radiation shields on the pattern as measured with a PbTe detector. The pattern of Figure 22-9a was obtained using a shield whose internal reflectivity was approximately zero and whose length was one half of its diameter. Here the intensity of 31 degrees has dropped to one half of its maximum value. Figure 22-9b shows the resulting pattern
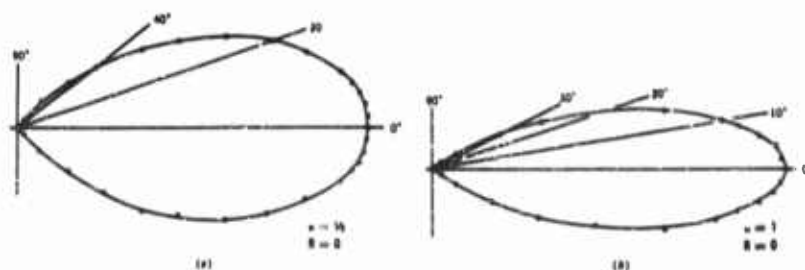
FIGURE 22-9   Radiation patterns of jet simulator
showing effects of shields (Reference 15)

for a shield whose internal reflectivity was approximately zero and whose
length was equal to its diameter. Here the intensity at 20 degrees has dropped
to one half its maximal value. In both cases the temperature of the shield
was kept sufficiently low so that it did not radiate appreciably. These plots
show that an appreciable narrowing of the radiation pattern occurs. Although
the measurements shown in Figure 22-9 were made using a PbTe cell, the
effectiveness of the shields in the PbS region would be at least as great since
the PbS cell would "see" a smaller fraction of the plume radiation. As
mentioned before, the plume radiation is small.

If such a shield were to be used on an aircraft, some sort of forced cooling
would have to be used on the outside of the shield to carry away the heat
absorbed by the black inside wall. This could be accomplished by having a
concentric outer cylinder surrounding the shield with air being forced
through the space between. This air could either be ram air picked up by
scoops at the front of the engine or air which is obtained from the compres-
sor. It has been estimated that the use of such a system, capable of keeping
the wall temperature at about 300 C would cause a performance penalty of
only about 3 percent (Reference 16) in the range of the aircraft. If the
shield were at a temperature of 300 C, the peak of the blackbody radiation
originating from it would be approximately 5 microns. This would probably
be cool enough for missiles with lead sulfide seekers. For missiles with longer
wavelength seekers some additional cooling would probably be desirable.

The use of radiation shields as an infrared countermeasure technique offers
several attractive features. Among these are: 1) The shield provides contin-
uous protection and does not have to be "turned on" when an attack is im-
minent. 2) The presence of the shield forces the enemy to launch his missiles
at smaller angles to the target aircraft axis. An especially important advan-
tage in this connection is that the use of shields minimizes the danger of a
"snap up" attack (i.e. the firing of a missile from an altitude considerably

less than that of the target bomber), thus requiring the enemy to expend more fuel and time in gaining an altitude from which the attack can be made. 3) The aerodynamic penalty caused by the shield and its cooling system is not excessive. 4) The greatly narrowed cone of radiation aft of the aircraft gives it a chance to avoid the missile by an evasive maneuver.

22.8 The Use of Smoke Clouds as an Infrared Countermeasure for Aircraft

The use of smoke dispensed from an aircraft has been proposed as an infrared countermeasure. Considerable experimental work has been done on this subject (Reference 17). Although it has been shown that smoke puffs and clouds can cause an infrared seeker, following an aircraft, to break lock, the proper conditions for the use of smokes are difficult to realize. First, to be effective the smoke must fill a reasonably large solid angle behind the aircraft, say 20° to 30°; secondly, the smoke must be of carefully controlled particle size to have its maximum scattering effectiveness in the proper spectral region and must also be of sufficient density. Smokes dispensed from aircraft tend to trail out in a very narrow beam, providing no protection against missiles launched outside of this narrow angle. Due to the relatively broad radiation pattern of an aircraft and scattering of this radiation by a narrow cone of smoke behind it, it is possible that the radiation reaching the missile may actually be increased, making the aircraft more vulnerable. Particle size is also difficult to control. This means that much of the smoke produced will be an ineffective scatterer.

The weight of chemicals which have to be carried on board the aircraft to produce an effective smoke cloud, filling a sufficient volume (assuming such a cloud could be produced), is probably far greater than the weight of pyrotechnic flares required to give an equivalent amount of protection.

The use of smoke puffs as scatterers of sunlight to decoy missiles has also been proposed (Reference 19). The difficulties involved in the use of screening smoke clouds discussed above also apply here. Other shortcomings of this technique are; 1) the radiation scattered from the smoke cloud would be largely rejected by the seeker due to its space filtering and spectral filtering features; and 2) this scheme could be used only in the daytime.

22.9 Infrared Countermeasures for Ground Installations

The protection of ground installation against surveillance or attack by infrared controlled devices involves problems which are quite different from those involved in the protection of aircraft. As mentioned in Chapter 21, ground targets are usually at lower temperatures than aircraft targets and therefore longer infrared wavelengths are involved. Also, ground targets are

usually spread out over a large area, where certain small regions (such as those containing blast furnaces or power plants) have very different radiation characteristics from the remainder of the area. In the case of ground targets, atmospheric attenuation and weather conditions are much more important factors than for the air-to-air situations.

Two methods which have been mentioned for the protection of ground installations against infrared devices are the use of decoy radiation sources and the use of camouflage techniques. Both have severe limitations. Since the ground target as a whole usually covers a considerable area and consists of various types of sources, it is difficult to simulate.

Camouflage techniques would involve the use of coverings or paints with controlled emissivities which would compensate for the temperature and emissivity differences of terrain and objects on the ground. By this reduction of contrast, detection would be made more difficult.

The most effective means for the protection of ground installations appears to be the use of screening agents. A considerable amount of development work has been carried out relative to the production and evaluation of smokes for this purpose. This technique also involves certain problems as the following analysis shows.

Von Mie and Blumer (References 19 and 20) have developed a theory for the scattering of radiation by simple particles for the case where the particle size and radiation wavelength are comparable. As long as it can be assumed that the particles of a cloud act independently, that the incident radiation is collimated, and that no scattered radiation reaches the detector, this theory can be used to compute the attenuation produced by a cloud of uniform sized particles. This leads to the following equation:

$$I/I_0 = \exp -k\pi a^2 nl \qquad (22\text{-}1)$$

where

$\quad I = $ transmitted radiation

$\quad I_0 = $ incident radiation

$\quad k = $ the scattering coefficient which is a function of the dielectric constant of the material and the ratio of particle radius to wavelength of the radiation

$\quad a = $ the particle radius

$\quad n = $ number of particles per unit volume

$\quad l = $ path length

It is convenient to eliminate the number of particles per unit volume, n, from Eq (22-1). This can be done in the following manner:

Let e_s be the mass per unit volume of the smoke cloud and e be the density

of the smoke material. Then the mass of a single particle will be $4\pi a^3 c/3$ and the number of particles per unit volume will be $3c_s/4\pi a^3 c$. Equation (22-1) now becomes:

$$I/I_0 = \exp -3kc_s l/4ac \qquad (22-2)$$

If c_s and l are kept constant, then for a given material, the transmitted fraction I/I_0 or its reciprocal, the attenuation factor, I_0/I is a function only of the scattering coefficient k and the particle radius a. The scattering coefficient itself is a function of the ratio of the particle radius to the wavelength. In
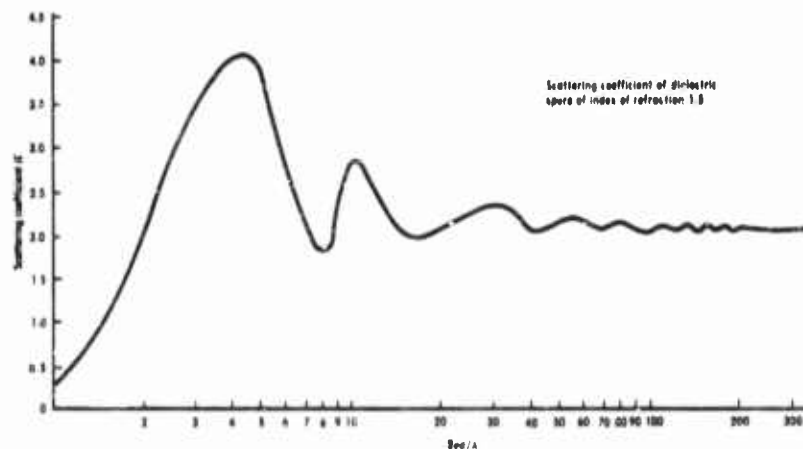
Figure 22-10   Scattering function, k plotted as a function of $2_\pi a/\lambda$

Figure 22-10, k is plotted as a function of $2\pi a/\lambda$ for particles whose dielectric constant is 1.5. It can be seen that successive maxima and minima occur as the ratio a/λ increases. Figure 22-11 shows the attenuation factor, I_0/I, plotted as a function of wavelength for particles of three different sizes. For the particles whose radius is 0.65 micron, the maximum attenuation factor is greater than 100,000 and this maximum occurs for radiation whose wavelength is approximately 1 micron. For radiation whose wavelength is 5 microns or beyond, the attenuation factor is almost zero. On the other hand, for particles whose radius is 2.6 microns, the maximum attenuation occurs at about 4 microns, but has a value of only about 20. These curves show, therefore, that as the wavelength increases, not only must the particle size incre. se but either the smoke density c_s, or the pathlength l (or both) must increase greatly if one is to maintain a constant attenuation factor. This analysis illustrates some of the difficulties involved when one attempts to use
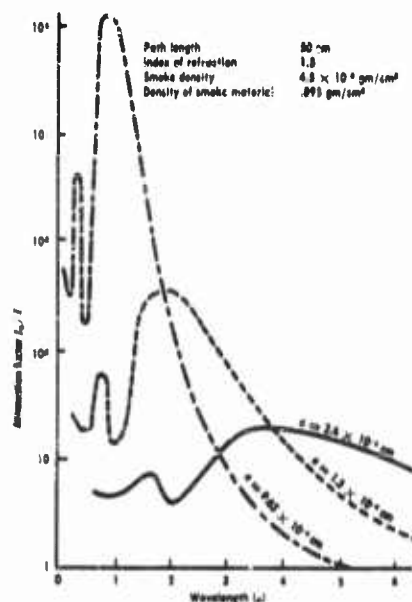
FIGURE 22-11 Attenuation factor versus
wavelength for clouds made up of three
different sized particles

a smoke cloud as an attenuation medium where long wavelength radiation is
involved. It shows that a smoke screen which is effective in screening a ground
installation against visual observation might be entirely ineffective for radia-
tion passing through the 4.5 to 5 and 8 to 13 micron atmospheric windows.
This was found from experiments to be the case (Reference 21). When field
tests and laboratory measurements were made on conventional smokes, such
as those produced from chlorosulfonic acid and iog oil, they were found to
become increasingly transparent beyond wavelengths of about 5 microns.
These tests and experiments confirmed the computed results, namely that
smokes of considerably larger particle size would be necessary for use at the
wavelengths involved in the radiation from ground targets. If conventional
materials were used, the settling rates of these larger particle size smokes
would probably be too great. One of the important requiremen's of a smoke
is that it stay suspended for a reasonably long period of time.

Experiments were performed in which plastic foams were produced in the
form of clouds of particles. One technique, which was successful, involved the
use of a gas turbine engine. The plastic resin and catalyst were sprayed

separately into the high temperature region of the gas stream. Separate (Reference 22) polymerized particles were formed which were dispersed into the atmosphere by the gas turbine engine exhaust. Measurements showed many of these particles to have diameters in the 20 to 25 micron range. Qualitatively, these smokes were found to cause considerable attenuation although it has not been possible to make quantitative measurements due to the difficulty of maintaining a constant pathlength in the smoke and to the fact that the density of the smoke cloud changes rapidly.

Although it does appear that it would be possible to provide an appreciable amount of protection for ground installations by smoke of this type, there are serious practical limita io s. The chief one of these is the problem of producing enough smoke to cover a large area, since the cloud tends to dissipate rapidly. Another problem is that ma v of these materials are of a toxic nature.

22.10 Engagement Warning System

For the most effective use of maneuvers and of any active countermeasures, i.e. one in which material is launched from the aircraft or otherwise consumed, some sort of warning or alarm system indicating that a missile has been launched is necessary. Warning systems are essentially passive search devices which operate by means of the radiation emitted by the rocket motor of the missile. The passive warning system is at a disadvantage, in that it must detect the approaching missile from the nose-on aspect. This is the direction in which the missile body and rocket nozzle provide a shield against most of the radiation. However, since the rocket plume extends aft of the missile and has a diameter somewhat larger than that of the missile, there is a substantial amount of radiation from the rocket in the forward aspect. Tables 21-IV and 21-V show the outputs of several typical air to air rockets in the 1.8 to 2.8 and 1.8 to 7 micron regions of the infrared and in the 2600Å to 2900Å region of the ultraviolet.

The ultraviolet radiation emitted by the rocket during the period that the rocket motor is burning might also provide a means for detecting the approach of the rocket by its intended target. The ozone layer (which has its maximum concentration around 80,000 feet, see Figure 22-12) fortunately provides a natural filter to remove most of the interfering solar background radiation when a wavelength of about 2550Å is used (see Figure 22-13). The problem of designing a detection system which is sufficiently sensitive at 2550Å, but which is completely insensitive on the long wavelength side of about 2800 or 2900Å is severe. Since the so-called "solar blind" detectors which have been developed have too large a sensitivity at the longer wavelengths, a suitable filter to be used in combination with the detector is necessary. Filters consisting of thin evaporated films of potassium enclosed
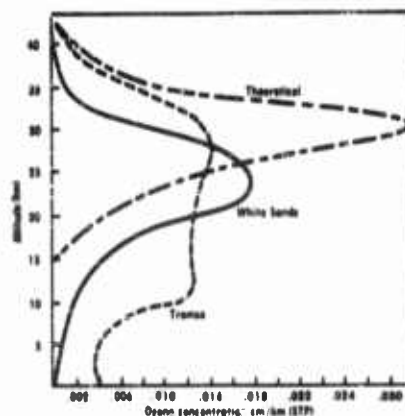
FIGURE 22-12 Ozone concentration as a
function of altitude (Reference 23)
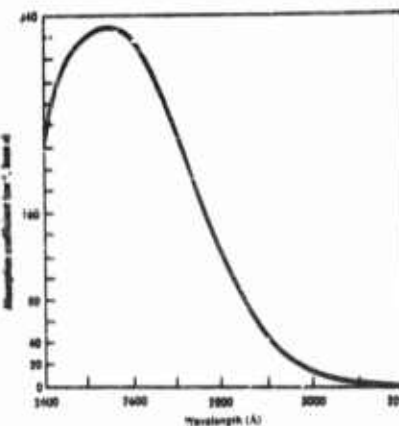
FIGURE 22-13 Absorption of ozone in the
ultraviolet spectral region (Reference 23)

between two quartz plates have been made. These filters transmit reasonably
well at the required wavelength but do not have a sufficiently sharp cut-off
toward longer wavelengths.

The principal drawbacks involved in the use of ultraviolet radiation in
engagement warning systems are listed below:

1) The best detectors which have been developed have too great a response
outside of the ozone absorption band.

2) The best filters available do not have a sufficiently large ratio of trans-
mission in the desired region to that at longer wavelengths.

3) At altitudes above about 40,000 feet the concentration of ozone in the
path between missile and detector becomes sufficiently great to cause severe
attenuation of the signal.

4) If ultraviolet wavelengths outside of the ozone absorption band were
used the scattered solar radiation would cause a very serious background
problem.

Several infrared systems for missile launch detection exist (Reference 24).
One of these consists of two search sets, one mounted at the top of the vertical
fin of the aircraft and one below the fuselage. Each search set consists of a
dome of infrared transparent material such as arsenic trisulfide glass con-
taining several detectors, each of which will observe a different sector in
elevation. The dome and the detectors rotate at a rate of a few revolutions
per second. If a signal is detected, its approximate elevation can be deter-
mined by noting the cell on which it was received; the azimuth can be

determined by the angular position of the scanner at the time of signal reception. One of the problems involved here is, of course, the difficulty of distinguishing the radiation of a launched rocket from the forward aspect radiation of the launching aircraft. Since the rocket exhaust is at a temperature considerably higher than that of the aircraft exhaust, the possibility of using a spect... filtering technique exists. The rocket radiation builds up and decays rapidly; thus there is also the possibility of using a time or pulse length discrimination technique.

Another system uses four detectors looking into the four quadrants aft of the aircraft. This system, upon the detection of a signal indicates only the quadrant from which the attack is being made.

Systems of this type will not indicate whether the object detected is an infrared guided weapon, a radar guided weapon, or an unguided rocket. The system simply indicates that countermeasures should be initiated. Since the time which elapses between the firing of an air-to-air missile and its reaching a position beyond which it cannot be deviated sufficiently from its path is very short, some sort of automatic device for the commencement of countermeasures activity will probably be necessary.

22.11 Integration of Infrared Devices With Other Electronic Countermeasures

There are several reasons why a completely integrated countermeasures system on an aircraft is desirable. Among these are: 1) A saving in power, space, and weight can be achieved since some of the auxiliary equipment required may serve more than one type of countermeasure. 2) In a penetration mission, it is likely that several types of countermeasures will be needed simultaneously. In an integrated system, a minimum of interference and confusion would resu!t ') Some techniques such as those used for flare and chaff dispersing may be quite similar. 4) With a single system, a minimum number of opera... g personnel in the aircraft would be required.

At present a development project is under way to develop a completely integrated countermeasures system for the B-52 aircraft (Reference 25). One of the infrared requirements specifies, for this system, flare batteries consisting of 51 flares each (probably 17 tubes containing three flares each). At present it is specified that each flare should have a 10-second burning time with an output of 3000 watts per steradian in the 1.8 to 2.8 micron region and 3000 watts per steradian output in the 3 to 5 micron region. The flares will probably be of the magnesium Teflon (Flora) type. An infrared warning receiver is also called for in this system. The complete specifications of this receiver are not available. It is specified, however, that the system noise of this device should be kept down to 1×10^{-11} watts per square centimeter in terms of equivalent radiation signal.

REFERENCES

1. Universal Match Corporation, Quarterly Progress Reports, Airborne Infrared Countermeasures, Contract AF 33(616)-3857. (SECRET)

2. Johns Hopkins University Radiation Laboratory, "Report on Symposium on Military Applications of Infrared Physics, January 25 and 26, 1956," Technical Report AF-31, Issued June 1956. (SECRET)

3. Universal Match Corporation, "Fifth Quarterly Interim Development Report on Airborne Infrared Countermeasures," Contract AF 33(616)-3857, April 25, 1958. (SECRET)

4. Hult, J. L., "Balls of Fire and Other Infrared Countermeasures," RAND Report RM 1353, October 25, 1954. (SECRET)

5. Armour Research Foundation of the Illinois Institute of Technology, Quarterly Progress Reports on Contract AF 33(616)-3596, 1957 and 1958. (SECRET)

6. Frese, R., and Y. Morita, "Final Report on Infrared Countermeasures," University of Michigan, Willow Run Laboratories, Report No. 2422-34-F, August 1958. (TOP SECRET)

7. Hallicrafters Company, "Final Engineering Report on Flare Dispensing Equipment," QRC 35(T) NTE 590-24-3-1. (SECRET)

8. Johns Hopkins University Radiation Laboratory, "Research Progress Report on Infrared Countermeasures," Quarterly Progress Report AF 3-2 Part II, July 1, 1956. (SECRET)

9. Burt, R. A., "REAC Investigation of the 'Blinker' Countermeasure to a Homing Missile," RAND Report RM 779, February 25, 1952. (SECRET)

10. Burt, R. A., and R. H. Frick, "REAC Simualtion of the 'Blinker' Countermeasure Against the Sidewinder and Falcon Missiles," RAND Report 822, April 28, 1952. (SECRET)

11. Burt, R. A., "The Three Dimensional Effect of the 'Blinker' Countermeasure on Homing Missiles," RAND Report RM 997, November 25, 1952. (SECRET)

12. Johns Hopkins University Radiation Laboratory, "Quarterly Report on Infrared Countermeasures," AF 2-15 Part II, Contract AF 33(616)-68, January 1, 1956. (SECRET)

13. Del Mar Engineering Laboratories, Final Engineering Report No. 541-7, Research on Infrared Aerial Towed Decoy System, July 1956. (SECRET)

14. Biberman, L. M., "A Review of Electro-Optical Countermeasures at NOTS," NAVORD Report 4984, November 30, 1955. (SECRET)

15. Snider, A. E., and F. Sobel, "Some Effects of Cylindrical Radiation Shields Part I," Johns Hopkins University Radiation Laboratory, Technical Report AF-29 Part I, March 1956. (SECRET)

16. Johns Hopkins University Radiation Laboratory, Internal Memorandum, RL/55/IMA-14, Report on Visit to England and France to obtain Information on Infrared Developments and Their Military Applications, July 1955. (SECRET)

17. Johns Hopkins University Radiation Laboratory, Quarterly Progress Report on Infrared Countermeasures, Contracts AF 33(616)-68 and AF 33(616)-3374. (SECRET)

18. Greenfield, S. M., "Preliminary Examination of Possible Infrared Countermeasures for Aircraft," RAND Report RM 962, October 15, 1952. (SECRET)

19. vonMie, Gustav, Beiträge Zur Optik trueben Medien, *Ann. Physik*, Vol. 25, p. 377, 1908. (UNCLASSIFIED)

20. Blumer, M., *Zeit. Physik*, Vol. 32, p. 119, 1925, and Vol. 38, p. 304, 1926. (UNCLAS-SIFIED)

21. Minkowski, J. M., and S. W. Gahagan, "The Effect of Particle Radius on the Scattering of Radiation by a Cloud of Spherical Particles." Johns Hopkins University Radiation Laboratory, Technical Report AF-11, September 1, 1954. (SECRET)

22. Johns Hopkins University Radiation Laboratory, "Research Progress Report on Infrared Countermeasures," Quarterly Progress Report AF 2-11, Contract AF 33(616)-68 January 1, 1955. (SECRET)

23. Johns Hopkins University Radiation Laboratory, "Quarterly Progress Reports on Infrared Countermeasures, Part II," AF 3-3, October 1, 1956. (SECRET)

24. Aerojet-General Corporation, Research on Airborne Infrared Countermeasures, Contract AF 33(600)-27010, Series of Quarterly Reports. (SECRET)

25. Sperry Gyroscope Company, Electronic Countermeasures System, AN/ALQ-27, Quarterly Reports, Contract AF 33(600)-35477. (SECRET)

23

Underwater Acoustic Countermeasures

R. T. McCANN

23.1 Introduction

The ocean presents a poor environment for the propagation of electromagnetic waves. Thus, underwater communication and detection must rely on other forms of energy propagation. Acoustic energy is most commonly used, in both the sonic and ultrasonic frequency ranges. For communications, the energy is either keyed, transmitted as a voice modulated carrier, or transmitted by some secure method. Detection is accomplished by either listening for noise emanating from the target or the reflection of an acoustic pulse by the target. The word SONAR is used to describe, in general, the detection and tracking of underwater targets, whether by passive listening or by actively transmitting and receiving pulses. Sonar systems are employed by aircraft, ships, and acoustic torpedoes to detect, track, and attack waterborne targets.

It is apparent that some means are necessary to deny an enemy the information required to press home an attack on our ships, either surface or subsurface. The most effective means would be to deny the enemy all possibility of detection; however, this is not feasible. It becomes necessary, therefore, to reduce the target to a minimum and thereby increase the detection problem for the enemy. Similarly, means must be provided to lessen his attack capability when detection is considered probable or certain. Submarines have an advantage over surface ships in that they can take advantage of the natural characteristics of the ocean environment as well as

their own ability to operate quietly for a limited time. Also, a submarine can generally detect a surface ship at greater range than the submarine can be detected by that ship. When the environment and tactical operation of the ship fail to provide adequate protection, a countermeasures system becomes vital to survival. Such a system must provide the knowledge that one is under surveillance, the means to break contact, and an alternative target or decoy.

23.2 Functions and Operational Concepts

A countermeasures system has three general functions: intercept, jamming or masking, and deception. These functions are normally performed in the same order as listed. Intercept occurs first since it is necessary to know that one is under surveillance and that further action may be necessary. It is usually desirable to break the sonar contact to force the enemy to renew the search and, also, to deny him a direct comparison between a true target and a false one. Finally, to escape or set the enemy up for counter-attack, it is desirable to provide a false target or decoy, acoustically similar to the true target.

The above sequence is more useful to a submarine than a surface ship since the latter is subject to visual and radar detection as well. Since such is the case, countermeasures in the form of maskers or decoys are frequently employed on a full time basis by surface ships when in an area of known or expected submarine activity. Since echo ranging would be carried out in such areas anyway, little, if any, advantage is lost by streaming an additional sound source.

Generally, it is considered undesirable for a submarine to make its presence known before an enemy has assured contact. On the other hand, too great a delay in taking countermeasures action may be fatal. Thus, the commanding officer of a submarine is faced with the dilemma of positively classifying himself or risking destruction. Only experience and confidence in the countermeasures system can develop the efficient employment of such a system. Where countermeasures are employed in offensive operations, such as penetrating a convoy screen, decoys should be launched from outside the detection range of the screen in such a manner as to cause a redeployment of the screening force. The submarine should find it easier to penetrate the weakened screen.

One approach to countering the enemy is to deny him the opportunity to launch a weapon. Thus, if a surface ship can detect and destroy a submarine beyond the submarine's weapon range capabilities, weapon countermeasures are not necessary. Similarly, a submarine that can deny a fire control solution to the enemy does not have to worry about the weapon. However,

optimism of this nature can be fatal and weapon countermeasures should
be a part of any such system. The first indication of an enemy's presence
might be the detection of a weapon. In such a case, the time for action is
short and the countermeasures system must be capable of immediate employ-
ment. Accordingly, surface ships frequently employ countermeasures on a
full-time basis for certain weapon types and rely upon the ability to inter-
pose countermeasures in time between the weapon and the ship for other
weapon types. Submarines, on the other hand, do not enjoy the privilege of
emitting much noise. Therefore, a submarine countermeasures system must
be in a standby condition such that it can be put into operation by pushing
a button, or even automatically, based upon signal level.

23.3 Functional Relationships

Although little has been done to develop an integrated acoustic counter-
measures system for either submarines or surface ships, such an approach
appears to have definite advantages over the individual equipment approach.
Surface ships, for example, rely on echo-ranging sonar to detect enemy sub-
marines and torpedoes. Any towed noisemaker increases the background level
through which the sonar operates, thus decreasing its detection range. Alter-
nate operation of the sonar and noisemaker partially solves that problem.
However, when a broadband receiver is added for intercept purposes and if
all equipments are to be used efficiently, the problem becomes quite complex.
Further complications arise when an expendable masker or a discrete fre-
quency jammer is placed in the water. A similar situation exists for the sub-
marine. Although the background noise problem is not so great, the submarine
must still detect and track the attacking ship and its weapons, take counter-
measures action, and escape.

Since nonacoustic methods of detection are presently unable to detect a
submerged submarine (although research being conducted in this area may
prove effective) and since the modern submarine is essentially a true sub-
mersible, more at home under water than on the surface, it is apparent that
acoustics provide the only ready approach to detecting and destroying sub-
marines. Submarines, therefore, should be primarily capable of countering
acoustic detection and attack.

The submarine platform will be used here as the basis for developing a
countermeasures system and showing the functional relationships of the sub-
systems. Figure 23-1 shows a functional block diagram of such a system.
The intercept subsystem provides the knowledge that the submarine is under
active sonar surveillance. The received sonar signals are analyzed and dis-
played by the next subsystem. This display also presents information from
the listening or passive sonar to permit the tracking of noisy targets. The
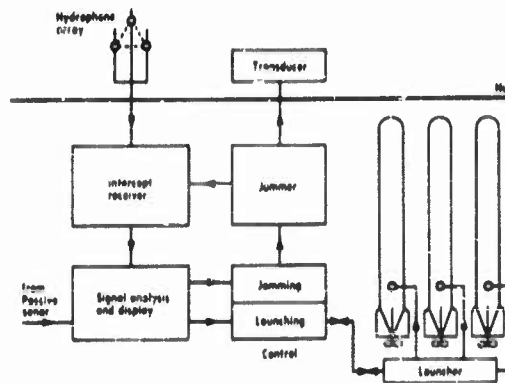
FIGURE 23-1   Countermeasures System Functional Block Diagram

analyzed signal data are fed to the controls of both the mounted and ex-
pendable jammers setting the jammers on frequency. In the presence of sig-
nals of different frequencies, the jammers could be set to transmit on a
time-shared basis, giving preference to the signal considered to be the greatest
threat. At such time as it is desirable to jam, the mounted jammer would be
activated. Upon activation, a signal feeds the intercept receiver causing
blanking of the portion of the frequency band being jammed. Thus, the
receiver can still accept signals in the remainder of the band. In the mean-
time, the expendable units are armed, programmed, and readied for launch-
ing. Launching readiness is relayed from the launcher to the launching con-
trol. Upon release, the expendable jammers commence transmitting and the
mounted jammer is secured. Since the jamming signal is the same from both
jammers and the location of the sources essentially the same, the switch
should go undetected.

The submarine is now free to maneuver to take advantage of the maneuver
program set into the jammer vehicle. One or more decoys would now be
launched from behind the jammer screen, giving the enemy a realistic sub-
marine target. This decoy target would be programmed, as was the jammer,
by the launching control. The program would be set to lead the enemy away
from the submarine such that the jammer screen remains between the sub-
marine and the enemy. Small hovering jammers and decoys can also be
launched to provide additional false targets should the enemy fail to classify
the decoy as a target.

23.4   Functional Integration

It is apparent that integration of the functions of the various components

is necessary to provide ready response to a threat with a minimum of equipment. Feeding the jammer from the receiver eliminates the need for a separate receiver associated with the mounted jammer. Similarly, a feed from the passive sonar eliminates the need for this function in the intercept receiver. The self-propelled vehicles would be designed to perform either the jamming or decoy function, reducing the number of vehicles carried and increasing their flexibility.

Intercept look-through, by blanking out the jammed portion of the frequency spectrum, has not been previously incorporated in either the hypothetical intercept receiver or the receiving section of the hypothetical jammer. The jammers have been designed with a receive-transmit cycling arrangement so that frequency shifts could be followed. However, there was a delay in following the shift equivalent to the transmitting time remaining in a particular cycle. The look-through feature in a countermeasures system will permit the following of frequency changes much more rapidly.

Some knowledge of the range to the enemy is necessary to intelligently make tactical decisions based upon the probability of having been detected. Presently an estimate is made based upon the intensity of his noise output and the repetition rate of his sonar. However, this does not provide accurate information. Passive ranging sets have been developed, but the complexity is such that their incorporation in the countermeasures system would be of questionable value, particularly since they operate on noise rather than sonar echoes (pings). A triangulation ranging system with a base line along the length of the submarine is limited in accuracy to short ranges normal, or nearly normal, to the base line. Its value is also questionable since sonar ranges have increased along with the ranges from which weapons can be delivered. Passive ranging underwater by the use of different signal paths may provide the solution, but insufficient work has been done in this area. Reasonable success has been achieved with a passive system using the paths in air of electromagnetic signals. No attempt has been made to incorporate a range solution in the system being hypothesized since it is felt that ranges to ships can be acquired by a passive fire control set or reasonably estimated by the aforementioned methods. Presently, the acquisition ranges of acoustic pinging torpedoes are such that immediate countermeasures action should be taken as soon as the weapon has been detected. The extravagance of having the option of deciding when to counter does not exist; it is a matter of necessity to act immediately. At such a time as torpedoes can detect targets at considerable range, accurate range information may become a necessity and warrant a complex set to perform the function.

The importance of timing in jammer use cannot be overemphasized. The time to start jamming is an on-the-spot tactical decision based on the serious-

ness of the threat; the time required to shift frequency and the relative time
assigned to each frequency, when more than one is being jammed, are design
considerations and are built into the equipment. Since jamming definitely
classifies the target as submarine, the time to start is the most critical of
the three and, due to the many variables involved, cannot be predicted with
any degree of accuracy other than on-the-spot. Thus, it is left as a command
decision. Although the time to shift frequency and the relative times assigned
to each of the frequencies to be jammed may be determined on-the-spot for
a mounted jammer, this is not the case for the expendable type. These fre-
quency shifts should be made automatically by the equipment. To assure
like transmissions, they should also be automatic in the mounted jammer,
perhaps with a manual override feature. Once launched, an expendable unit
normally can no longer be controlled from the submarine.

To be effective, a submarine decoy should provide doppler echoes, propul-
sion noise, wake simulation, and maneuvering characteristics comparable to
that of a submarine. Hovering decoys are rather simple and, generally, pro-
vide only a doppler echo. The amount of doppler shift is fixed for a given
frequency band and corresponds to that of an average speed submarine.
Self-propelled decoys, due to their movement, require no built-in doppler.
Propulsion noise corresponds to the gear whine and propeller beat of the
submarine and is representative of types of submarines rather than any given
one. The degree of realism must be general rather than specific. A target
with the general characteristics of a submarine is apt to be so classified, par-
ticularly when a direct comparison with the submarine is not possible. The
decoy life should be limited to the time required by the submarine to clear
the area. A decoy with a long life may permit the enemy to discover, by
prolonged tracking and analysis, a characteristic that is not present in a
submarine. From then on, this characteristic would be sought out initially
and obviate classification as a submarine.

23.5 Interception

The basic requisite of any countermeasures system is a means to provide
knowledge that one is under surveillance. The underwater intercept receiver
provides such information. The submarine will be used here as the platform
since its intercept requirements are more inclusive than those of a surface
ship. The basic difference between an intercept receiver and a passive sonar
is that the former is designed to intercept deliberately transmitted signals,
whereas the latter is designed to detect the noise ships make in transiting the
seas.

Early intercept was provided by passive sonar with a supersonic converter.
Such an equipment required manual scanning in both azimuth and frequency

with a resultant very low probability of intercept. Based upon the premise
that the bearing to any threat, assumed to be a waterborne vehicle, could
be determined by the passive sonar, the intercept design was based on omni-
directional, frequency scanning receivers of the panoramic type. High back-
ground noise levels rendered these receivers practically useless. Several re-
ceivers were developed following World War II using various techniques for
detecting discrete frequency signals within a broadband. The trend toward
passive sonars without supersonic converters coincided with the completion
of the above developments and in the absence of noise from the transmitting
ships, it was no longer possible to determine the source bearing of an echo-
ranging sonar. (The converter extends the frequency range of the passive
sonar receiver, enabling it to detect echo-ranging sonars.) The requirement
for providing bearing data integral with frequency information was quite
evident.

The first significant effort in this direction resulted in a receiver system
made up of a spherical lens hydrophone and a correlation type receiver.

FIGURE 23-2   Underwater Intercept Receiver, Q-8p, Block Diagram
(Courtesy U. S. Navy Electronics Laboratory.)
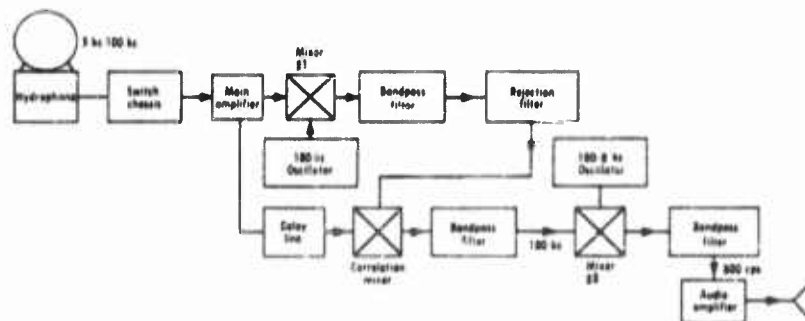
Figure 23-2 presents a block diagram of the equipment developed. The
hydrophone was made of 36 elements around the equator of a sphere which
was filled with a focusing fluid. An incoming signal would be focused on the
element on the side of the sphere opposite to the transmitting source. Three
modes of operation were provided by which one could observe 360 degrees
of azimuth, a 60 degrees sector or a 10 degrees sector. The equipment was
normally operated for 360 degrees coverage. Upon detection of a signal, a
push-button keyboard permitted localization to the proper 60 degrees sector.
Switching to the 10 degrees sector mode of operation, the same keyboard
permitted one to determine the element receiving the highest signal level
and thus the bearing to a 10 degrees sector. However, such an arrangement

required time to narrow down the bearing. A wave trap filter was scanned across the frequency band to determine the signal frequency. The main weakness of this system became apparent if the source ceased to transmit prior to receiving enough pulses to determine both bearing and frequency. A trained operator could usually obtain the necessary data from 10 to 12 pings. For a source at 5000 yards, the data could usually be obtained in no less than one minute. Should an echo-ranging torpedo be launched within 1000 yards of the submarine, insufficient time would remain for countermeasures action. The incoming signal (see Figure 23-2 again) was amplified and split such that part was delayed and the remainder heterodyned and filtered to pass the upper sideband. The delayed and heterod'... signals were mixed resulting in the oscillator frequency output. This output was again mixed with another signal from an oscillator 800 cycles per second higher than the first, resulting in an 800 cps tone which was filtered, amplified and fed to a speaker or headset. Only signals of duration longer than the delay were correlated into an output. Short pulses, such as noise, resulted in no output from the correlation mixer.

To overcome the deficiencies of the first correlation receiver, a system was developed which would automatically indicate the bearing and frequency of a transmitted pulse. This system employed three hydrophones, placed at the vertices of an equilateral triangle from which the bearing was determined by the difference in time of arrival of the signal at the hydrophones. At first, correlation, similar to that described above, was employed only in the detection channel. However, noise pulses at the hydrophones tended to throw the bearing off, particularly at higher speeds. Further development resulted in placing the correlation delay in all three bearing channels; false bearing readings are reduced considerably. Figure 23-3 presents a block diagram of the improved system for the intercept receiver. The operation of the correlation circuits in the bearing channels is the same as described above. One of these channels also feeds the audio output. The source bearing is determined from the trigonometric relationships between the equilateral triangle formed by the hydrophone array and the incident wave. The distance between hydrophones traveled by the wave divided by the velocity of sound gives the time the wave travels between the hydrophones. The differences in times of arrival are measured in the bearing circuits of the hydrophones. These time differences are transformed into an a-c signal and fed to the windings of a bearing synchro. The combination of these signals in the synchro results in the shaft displacement, and the source bearing is read directly from the meter which is calibrated in degrees of relative bearing.

Essentially, a cycle counter is used to determine frequency. Each cycle generates a fixed increment of internal signal. These increments are added
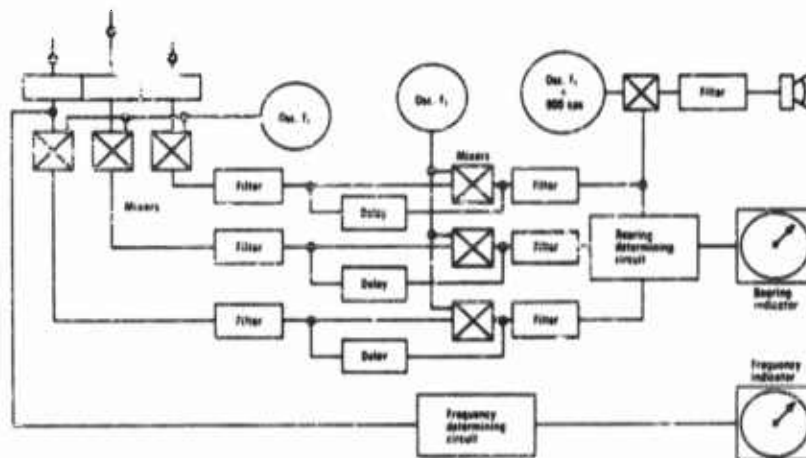
FIGURE 23-3   Underwater Intercept Receiver, AN/WLR-2, Equipment Block Diagram
(Courtesy General Electric Co.)

and fed to the meter which is calibrated in frequency. Thus, the intercept
receiver can indicate, from a single ping, frequency and source bearing. An
alarm is incorporated to call the attention of an operator should the equip-
ment be in unattended operation. A hold feature retains the meter readings
for a fixed period of time or until reset.

An intercept receiver must cover a frequency band which includes all
known or anticipated signals of a threatening nature. One must be able to
detect all echo ranging torpedoes, which generally operate at higher fre-
quencies and set the upper limits of the band, as well as the lower frequency
search sonars. Both torpedo and sonar designers tend to take advantage of
the greater ranges possible at lower frequencies. During World War II,
torpedoes operated in the 60 to 80 kilocycle per second frequency range,
whereas sonars were pretty much confined to the 20 to 40 kilocycle per
second frequency range. In the near future, it is quite possible that search
sonars will operate below 15 kcps and torpedoes between 20 and 40 kcps.
Since the sonars will operate at frequencies as low as a few hundred cycles
per second, the intercept problem should become progressively more difficult.
If the frequency range from 1 to 80 kcps is effectively covered, one can
presently expect to detect just about all the operational sonars and weapons.
However, in a few years, this coverage will be extended to include a band
from about 50 to 100 cps at the low end and up to 80 kcps at the high end,
to be reasonably certain of adequate intercept capability.

The pulse response of an intercept receiver is also a vital factor to be

considered. It is necessary to have the capability of detecting and analyzing
pulsed signals from as short a time as one millisecond to a few hundred
milliseconds and even continuous wave signals. It is not likely that pulses
will become shorter than one millisecond since it is still necessary to get the
signal information, and, at the same time, discriminate against noise bursts.
In general, torpedoes employ short ping lengths of a few milliseconds, whereas
sonars use longer pings.

It is difficult to define the sensitivity of an intercept receiver in a practical
manner. If one specifies high sensitivity, the false alarm rate during high
noise background conditions is intolerable. A low sensitivity can reduce an
intercept receiver's range to that equal to or less than the ranges obtained
under low noise conditions with a sonar. It is necessary to have a threshold
level control such that the intercept receiver always has a detection range
advantage over the sonar. Since the same environmental factors affect both
the sonar and intercept receiver, the sensitivity can be expressed as a func-
tion of sonar detection range. Sonar detection range is generally defined as
that range at which there is a 50 percent probability of detection under the
existing conditions. If one specifies the intercept sensitivity as being an 85
to 90 percent probability of detecting the sonar signal at the sonar detection
range, for the existing conditions, the sonar will rarely receive an echo from
the submarine carrying the intercept gear if it chooses to take evasive action.

One weakness of all intercept receivers developed to date has been the
lack of jammer discrimination. To make effective use of such a receiver in
a jamming environment, it must be capable of reducing the jammer influence
without reducing the sensitivity across the entire frequency band. To be
most effective with a jammer, one must fill the effective bandwidth that a
sonar can accept. Likewise, this same bandwidth must be rejected by the
intercept receiver. Since sonar receiver bandwidths vary with frequency and
pulsewidth, a suitable jammer rejection filter in an intercept receiver must
have variable bandwidth and be tunable across the frequency band being
covered. An alternative would be the use of a series of adjacent filters across
the band that could be inserted one or more at a time. Such an arrange-
ment would be less efficient since, to keep the number of filters to a practical
minimum, the rejected bandwidth would be more than the jammer band-
width. Since jammers have outputs of 98 to 100 db//1 dyne/cm² (corrected
to 1 yard), and because of the proximity of the jammer and intercept receiver,
the rejection of the filter must be sizable. The rejection requirement is not
so great with expendable jammers since the receiver to jammer spacing is
greater and the jammer output is 5 to 6 decibels lower.

The primary factor which generates confidence in an intercept receiver is
its probability of intercept with a minimum of false alarms. Confidence is

soon lost if one cannot rely on picking up sonar signals prior to being de-tected by those signals. On the other hand, great reliance will be placed in an equipment which has proven that it can permit a submarine to maneuver beyond sonar range until such time as it is desired to do otherwise. Scan-ning, both frequency and azimuth, requires time which reduces intercept probability, particularly against short signals such as a torpedo transmits. Thus, the most dangerous signals are most apt to be missed in a scanning system. A receiver that is "wide-open" in both azimuth and frequency is required to overcome this liability. To be sure, such equipment is not as sensitive as a scanning receiver but, for tactical purposes, one is not too concerned about signals of very low level since these will not return echoes to the sonar.

Intelligence type receivers have not been mentioned because the require-ments are different. The probability of intercept of an intelligence type re-ceiver need not be as high as that of a tactical equipment. However, a high sensitivity and good signal analysis capability are necessary to a higher degree than in the tactical receiver.

23.6 Jamming

Jamming is a vital function to be performed by any countermeasures system. The purpose of jamming is to break the sonar contact causing the enemy to begin a new search for the target. It is most difficult to "sell" a decoy if contact is maintained with the real target. Thus, breaking the con-tact by jamming actually hides the real target and creates a situation in which the decoy is more likely to be classified as a target. The relative merits of masking and jamming have been argued many times. Each method of breaking contact has its merits. Against passive sonar, masking is the only suitable method since it uses a broadband noise source and there is no good way to determine the acceptance band-width of the sonar. Against active sonar, masking has little to recommend it, ex-cept against very short pulses which are in the order of duration of noise pulses. Jamming centered on the sonar frequency with a bandwidth approximating that of the sonar is much more effective for pulses of practical length. Figure 23-4 graphically presents the relative effective-ness of a noisemaker or masker, and a jammer against very short-pulse sonar. It is quite apparent that the jammer is
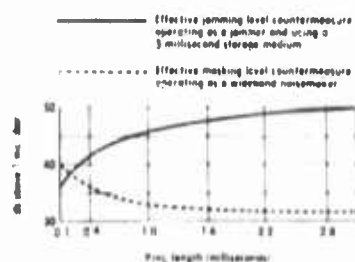
FIGURE 23-4 Relative Effectiveness of a Noisemaker and a Jammer When Used Against Very Short-Pulse Sonar

more effective for pulse lengths in excess of about 0.2 milliseconds.

Initial attempts at jammer development resulted in canister and self-propelled jammers which had receive and transmit modes of operation using a single transducer. The cycle required 30 seconds to complete and consisted of a 5 second receive and 25 second transmit mode of operation. When a signal was received, the receive mode was automatically switched to the transmit mode resulting in a transmission of 25 seconds plus the remainder of the 5 second receive mode. Following transmission, there was a brief dead period to permit reverberations to die out before another reception. This precluded triggering on the previous jammer transmission. In the event that no new pulses were received within the 5 seconds, the previous signal was retransmitted. Thus, changes in the sonar frequency could be followed with a delay no greater than 30 seconds. The active life of such jammers was on the order of 20 to 30 minutes and was limited primarily by battery capacity. A similar jammer was developed for mounting on submarines. This equipment had greater output and a manual override feature. The override feature permitted recycling to the receive mode at any time and was most useful when confronted with a multiple sonar attack. Should the jammer be transmitting on the sonar frequency considered less threatening, it could be recycled until it transmitted on the most threatening sonar frequency. All of these jammers give statistical preference to the sonar with the highest repetition rate since the probability of receiving that sonar is greater.

In a later development the masking and jamming features were combined
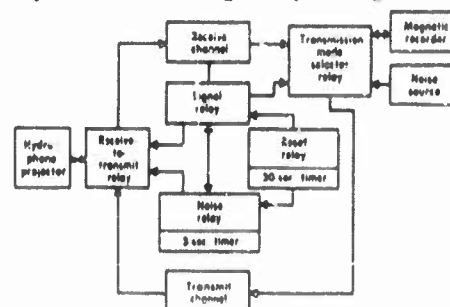
FIGURE 23-5   NAH Canister Beacon, Block Diagram
(Courtesy U. S. Navy Electronics Laboratory.)

in a single equipment. Figure 23-5 shows a block diagram of the beacon which transmits both jamming and masking signals on a time-shared basis. If no signal is received, broadband noise is transmitted for 25 seconds. If a signal is received in the listening mode, the operation is the same as de-
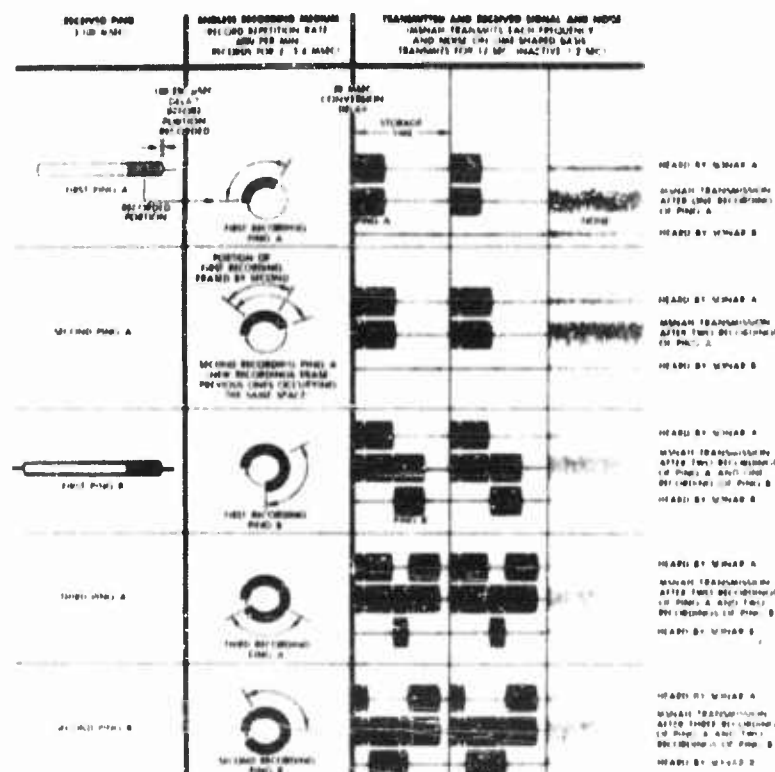
FIGURE 23-6   Character of Multiple Signal Jammer
(Courtesy U. S. Navy Electronics Laboratory.)

scribed above for the straight jammer. The latest development, known as
the multiple signal jammer (MSNAH), time-shares noise with several sonar
frequencies. Figure 23-6 indicates the character of the jammer transmission
after the reception of several pings of two different sonar frequencies. This
equipment permits jamming of more than one sonar as long as the recorded
frequency samples are relatively short and repeated at short intervals. If
the spacing between jamming pulses at any given sonar receiver is greater
than the retentivity of the ear or the persistence of the sonar cathode-ray
tube, some effectiveness is lost.

It has been determined that a jammer that can be programmed to travel
a given course is more effective than a stationary or hovering jammer whose
output may be 10 decibels greater. This is due to the fact that the submarine,

having set the program into the jammer, can maneuver to take best advantage of the jamming cover. The requirements on a jammer are such that it is nearly impossible to provide effective cover throughout the entire attack run of a surface ship planning to use depth charges. Generally, the submarine echo starts to show through the jamming at a range of 200 to 300 yards to the ship. However in a fast closing attack the attacker has gone beyond the point at which a live control solution can be developed for anything and depth charges launched from the stern. Even these would be relatively ineffective unless the attack run turned out to be passing directly over the submarine target.

The effectiveness of jamming against sonar is demonstrated in Figures 23-7, 23-8, and 23-9 which are photographs of a sonar presentation following three successive pings. Figure 23-7 shows a target clearly at about 340 degrees and about three-fourths of an inch from the end of the bearing cursor. The main body of the target appears to the left of the cursor. Figure 23-8 shows the same presentation after the next ping with intense brightening in the target area and general brightening over the entire scope. It is impossible to determine the range to the target and the bearing can only be approximated by bisecting the sector of greatest jamming intensity. Figure 23-9 shows the presentation after the next ping, the gain having been considerably reduced (estimated 20 to 30 decibels). There is still no trace of the target through the jamming. Under this condition of reduced gain, the target would be difficult to detect even if it should emerge from the jamming wedge. Figure 23-10 shows a submarine emerging from a jamming wedge at close range, estimated at 600 to 700 yards. It is apparent that the target would not be discernible if it had remained in the jammed sector. The jammer is about 500 yards from the sonar, and the bearings to the jammer and submarine differ by about 30 degrees. The limitation of the jammer is readily seen.

The jamming signal should be a fair reproduction of the received sonar signal. The recorded sonar signal will be shifted slightly from the transmitted one due to the relative motion between the ship and submarine. In addition, frequency modulation caused by slight speed variations in the recorder motor further alter the signal. In the event that the pulse length is greater than the time required for one revolution of the recorder disc, an overlap occurs resulting in a phase discontinuity at the overlap. Thus, there are enough uncontrollable factors affecting the quality of reproduction and care should be taken not to introduce other causes for signal deterioration.

The receiving pattern of the hydrophone should, for a mounted jammer at least, closely resemble the target strength pattern of a submarine. If such is the case, and the jammer is set to operate on threshold signals, it will be more likely to trigger on only those signals which would return a recogniz-

able echo. Thus, the receiving sensitivity is tailored to a typical submarine target strength pattern. Since a submarine presents a stronger target at the beams, compared to bow or stern, the jammer is more sensitive in these areas and less in the fore and aft directions. Whether the transducer maintains the same pattern during transmission is not too important since the submarine's presence is known anyway.

The most effective material for jammer and decoy transducers has been X-cut Rochelle salt. Although the temperature characteristics of Rochelle salt are well known, this drawback is overshadowed by its over-all sensitivity and power handling capabilities over a broad frequency band. Other transducer materials may be better for either receiving or transmitting of both over a narrow band, but for countermeasures requirements, Rochelle salt has proven best. The narrow hole in the frequency response of Rochelle salt would be completely unacceptable for equipment operating at a given frequency all the time. Since the frequency at which this hole occurs is fairly unpredictable, and since a jammer or decoy operates over such a wide frequency band, the probability of the hole occurring at the particular frequency to be used at any given time is rather remote. Little difficulty has been experienced with melting of the Rochelle salt due to exposure to high temperatures. Even the topside mounted transducer of a mounted jammer introduced no problem. However, an unpainted dome of stainless steel was used to reflect the rays of the sun.

To be effective against present day sonars, a jammer that is mounted on a submarine should have an output on the order of 100 db//1 dyne/cm^2 (corrected to 1 yard). A jammer that is self-propelled can get away with less output since it can be interposed between the sonar and the target submarine. The primary consideration is how much of the jammer output is useful in hiding the target. In general, the jamming signal should be as strong as the target echo after processing in the enemy's sonar receiver. The sonar's directivity will reduce the effectiveness of the jammer if there is a difference in bearing to the target and the jammer. It cannot be expected that a target will be hidden if its bearing differs from that of the jammer by more than about 20 degrees. In the case where the jammer is considerably closer to the sonar than the target, the bearing difference can be greater, possibly 40 degrees. Where the jamming bandwidth is greater than the sonar acceptance band, only that portion of the jamming signal within the acceptance band will be effective. Echo-ranging sonars may have acceptance bandwidths of 300 cps to a maximum of about 1000 cps. The above mentioned jammer output level should be for a 300 cps bandwidth, rather than a wideband level. Wideband noise is generally less effective against echo-ranging sonars since most of the energy falls outside the sonar acceptance band.
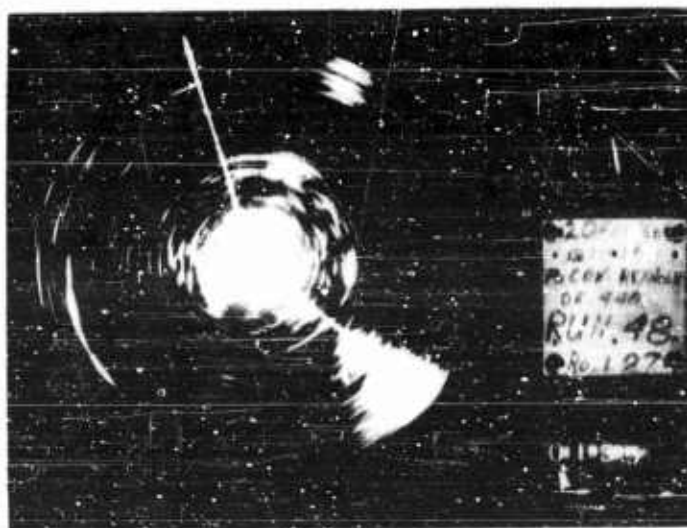
FIGURE 23-7   Sonar Presentation—No Jamming. Cursor on target at 340°.
(Courtesy U. S. Navy Electronics Laboratory.)
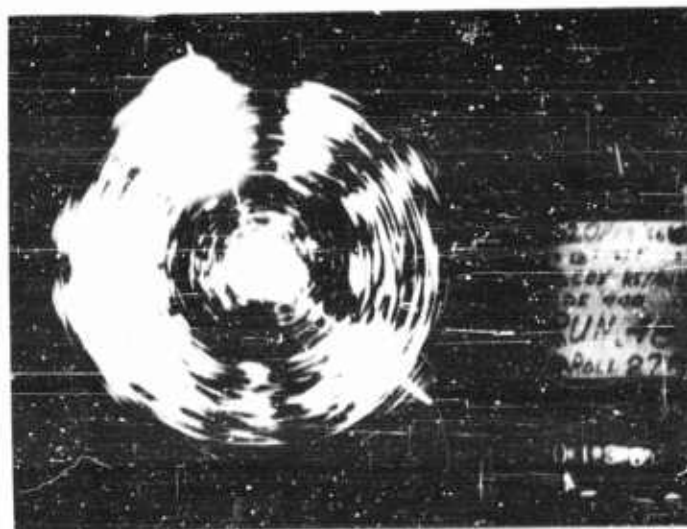
FIGURE 23-8   Sonar Presentation—Jamming Present—Normal Gain Setting on Sonar
(Courtesy U. S. Navy Electronics Laboratory.)

FIGURE 23-9 Sonar Presentation—Jamming Present—Reduced Gain on Sonar
(Courtesy U. S. Navy Electronics Laboratory.)

FIGURE 23-10 Submarine Target Emerging From Jamming Sector.
Range Scale 1500 Yards
(Courtesy U. S. Navy Electronics Laboratory.)

A prime consideration in jammer design is the compatibility of the signals from different jammer types. The signal characteristics should be as similar as possible so that they appear alike to echo-ranging sonars. Thus, when a mounted jammer is secured and a mobile or stationary type takes over, no difference in the signals should be detectable by the sonar.

23.7 Deception

Assuming a submarine can successfully detect echo-ranging sonars and jam them, it can still be held down in an area with little chance for escape. The battle then becomes one of endurance in which the submarine is hard pressed to compete with the surface, subsurface, and air forces which can be employed as the battle continues. If, however, the submarine can provide, early in the game, a substitute target or decoy with sufficient realism to attract the attackers, then there is a much greater probability of escape. Thus, the need for deception devices arises.

The simulation of a submarine does not have to be perfect by any means. A decoy which returns an echo similar to that from a submarine, emits sounds like a submarine, provides a wake comparable to that of a submarine, and maneuvers within the limits of a submarine's capability, has a good chance of being classified as a submarine. This classification is less assured if a direct comparison between the decoy and submarine is available to the enemy. If the sonar contact with the submarine can be broken by jamming, and a decoy launched so as to emerge from behind the jamming screen, the probability of its being classified as a submarine is good. Tests conducted in late 1955 and early 1956 with developmental countermeasures equipments in various combinations indicated that a submarine improves his escape probability by a factor of about five when using countermeasures compared to not using countermeasures. This is based on a total of 61 runs and 99 simulated attacks by surface ships operating singly and in pairs.

The quality of reproduction of a submarine's characteristics can be representative of a general class. Thus, fleet submarines should have one type decoy, guppy submarines another and nuclear submarines still another. However, this does not mean that several completely different decoys are required since the type of propulsion noise radiated is the primary difference in acoustic characteristics. A vehicle designed to simulate the maneuvering of a high speed nuclear submarine can be slowed down, its turn radius altered, and its operating depth stratum varied by programming. The echo characteristics may have to be altered but probably not beyond the adjustment limits of a well designed echo repeater. The radiated noise output, however, may be a more critical factor that requires a separate source in each case.

However, it is a simple matter to interchange noise simulators to meet specific needs.

The simplest decoy simulates a single submarine characteristic rather than all of them. A dopplerized echo repeater, for example, is a hovering type, has no wake or propulsion noise simulation, but just returns an echo
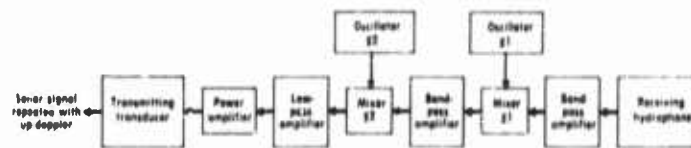
FIGURE 23-11   Hovering Decoy Block Diagram. *Note:* Oscillator #2 is slightly lower in frequency than Oscillator #1. For the lower-band decoys the difference in frequencies between the two oscillators is 60 ± 30 cps. For the higher-band decoys the difference is 100 ± 50 cps.

which has reasonable doppler characteristics. Figure 23-11 is a block diagram of such a decoy. The sonar pulse is received and amplified, heterodyned up and then down, amplified again and retransmitted. It can readily be seen that, if oscillator #1 is slightly higher in frequency than oscillator #2, the output signal will be higher in frequency than the received signal by the difference between the two oscillator frequencies. If this frequency difference is about 60 cps and 100 cps for the lower and higher band echo repeaters respectively, the doppler characteristic simulates a low-speed submarine. (A knowledge of the echo-ranging frequency, as provided by an intercept receiver, dictates which band echo repeater to employ.) Since a low-speed submarine, in general, radiates less noise than a high speed one, the absence of radiated noise is not too significant at the limit of the sonar detection range. At shorter ranges, the fact that it is a decoy is more apparent.

Figure 23-12 is a functional block diagram of a self-propelled submarine decoy. Upon launching, it follows a preset programmed course generating a
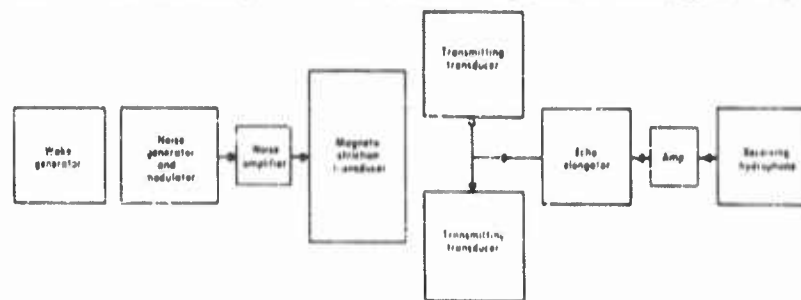
FIGURE 23-12   Functional Block Diagram of Mobile Decoy

wake and transmitting modulated noise which simulates propulsion noise. At such time as it receives a ping from an echo-ranging sonar, it repeats that ping in the form of an echo transmitted back to the sonar. It may or may not employ echo-elongation which indicates to the sonar the submarine aspect by the length of the echo. The target strength is similar to that of a submarine since the transducers are side mounted and the vehicle itself provides shielding in the fore and aft directions. The sonic simulation of propulsion noise is accomplished by modulating a noise generator such as to present a signal representative of propeller beats superimposed on machinery and gear noise. This signal is radiated by a magnetostriction transducer wrapped around, or integral with, the decoy hull. The wake is developed by the chemical reaction of lithium hydride with sea water. The lithium hydride is in the form of small balls, or globules, of various sizes coated with a water soluble substance. The ball sizes are such that when the reaction produces the gas, the resultant bubbles vary in size so that there are a large number resonant at all frequencies within the echo-ranging frequency band. A decoy represented by the diagram of Figure 23-12 is effective against echo-ranging sonars, passive sonars, and wake following torpedoes.

There are natural countermeasures available to the submarine. The natural temperature gradient in the ocean often provides a thermocline which, if sharp enough, is nearly impenetrable by echo-ranging sonars. If the sonar source is above the thermocline, a submarine can employ it to hide. Whales and schools of fish often return echoes mistaken for a submarine echo. A rapid turn or backing down by the submarine results in a concentrated wake, or knuckle, which is often taken for the submarine itself. False knuckle can also be created by chemicals.

Jammers can be considered decoys to a certain degree. If the enemy concludes that the jammer is aboard the submarine and attacks as best as he can by dead reckoning on the last target position and follows any bearing drift of the jamming wedge, he is effectively decoyed when the submarine launches an expendable jammer and secures the mounted one as he opens range. On the other hand, if the enemy concludes that the jammer is not on the submarine, he is not likely to attack the source of jamming.

Figure 23-13 depicts simultaneously three target echoes on a sonar presentation. The top echo, upon which the bearing cursor is trained, is the submarine; below that at about 270 degrees is a turn knuckle created by the submarine; still farther below at about 225 degrees is an AN/SQQ-9 decoy.

Target aspect has been mentioned only briefly. However, with greater refinements of sonars, it may become a more important feature in future decoys. It is well known that the sonar return from a submarine is not a single pulse but a series of adjacent short pulses caused by multiple reflec-

FIGURE 23-13 Sonar Presentation Showing Three Targets
(*Courtesy U. S. Navy Electronics Laboratory.*)

tions from the outer hull, tanks, superstructure, etc. Thus, the echo length is a function of the target length along the line of bearing from the sonar to the target. To return an echo with aspect from a decoy, it becomes necessary to determine the bearing of the sonar source relative to the decoy. However, it is not necessary to know this bearing throughout 360 degrees. It is sufficient to determine the angle of bearing relative to the longitudinal axis of the decoy since the echo length will be the same for a given angle, whether to port or starboard. The bearing angle can be determined using the difference in time of arrival of a pulse at two hydrophones on the decoy. The length of the echo repeated can then be controlled from a minimum for the beam to maxima for the bow and stern aspects.

Two types of decoys appear to have merit but have received little or no emphasis. The first is the strategic decoy and the second a weapon decoy. The strategic decoy is a long range device which simulates a submarine and would be, in fact, a small unmanned submarine. Such decoys would be used to dilute an enemy's ASW surveillance effort and create the impression that many more submarines are operating in a given area than is actually the case. By so doing, the forces available for any given contact, whether against a submarine or a decoy, would be reduced. An intermediate range strategic decoy would be launched several miles from a submarine's operating area,

travel to an area 50 to 100 miles or more from the submarine area, and maneuver there. Several such decoys launched from beyond the enemy's detection range could be programmed so that all the decoys and the submarine enter the surveillance area at about the same time over a front a few hundred miles long. If the decoys are realistic enough, the submarine should be able to carry out its mission with a much higher probability of success. A long range strategic decoy would do essentially the same thing, but be launched from advance bases, surface ships, or even large aircraft. Upon completion of its mission, the decoy would be scuttled as are the tactical decoys.

A weapon decoy is one that appears like a weapon, for example, a torpedo or a weapon that appears to be a countermeasure. Thus, a submarine could launch a torpedo and several inexpensive devices that appear to be torpedoes to the detection equipment. The enemy would be hard pressed to determine which to avoid or counter. Similarly, an active torpedo could have a built-in jammer with a slot cut out of its frequency spectrum to echo-range through, or a passive torpedo could have a slot cut out to listen through. When these torpedoes are echo-ranged on, they would jam the sonar, reducing its detection capability. Not only would this reduce countermeasures action against the torpedo, but it would also create a healthy respect for jammers since they would be known to "bite back" on occasion. Running down a jamming spoke would then be a dangerous practice.

23.8 Vehicles

It is necessary to use towed or expendable type vehicles from submarines and surface ships. The requirements for such vehicles are considerably different for the two platforms.

From surface ships, the vehicle is usually employed against torpedoes and, as such, must be capable of immediate and rapid launching to a range beyond the destructive radius of the weapon. Rocket launching provides about the best means to effectively interpose a countermeasures device between the weapon and target. The problem essentially resolves into one of detection and evaluation of the threat, training the launcher, and firing. It can readily be seen that the countermeasures must be ready for firing at all times from a remote position.

Upon entering the water, the device must maintain itself in the same stratum as the oncoming weapon. In addition, the transducer must be at a reasonable depth for good sound transmission and reception. Usually the transducer depth is maintained at about 40 or 50 feet beneath the surface. This is readily accomplished by a line between the transducer and a surface float; by means of an impeller driven by a reversible motor whose direction of rotation is determined by a hydrostatic pressure switch; or by means of a

gas bag arrangement whereby a chemical reaction produces gas to fill a bag which maintains the proper buoyancy for the desired depth.

Devices of the above types usually employ electronic or mechanical sound sources. The electronic devices are of the echo repeater variety from which a pinging torpedo receives an echo that is stronger than that from the target. By filtering, the bandwidth of the echo-ranging sonar can be eliminated from the spectrum, thereby reducing interference with the sonar. Masking devices are generally mechanical and emit continuous broadband noise. These are most effective against a torpedo which homes on target noise. The mechanical devices are quite rugged and consist of a motor driving hammers, balls, or rollers against the inside of a cylinder. The cylinder wall thickness, motor speed, and number of hammers are designed to produce the desired frequency spectrum. The electronic devices, on the other hand, are not so rugged and the water entry angle is more critical. The transducer is delicate and must be properly designed to withstand the shock of water entry. Since the force on the vehicle at launching is opposite from the force at water entry, the electronic tubes and other plug-in components should be mounted so that they tend to seat or reseat themselves on the impact of water entry.

Submarines employ vehicles of three basic configurations. Hovering vehicles are launched from the garbage ejection tube which is about ten inches in diameter, and the signal flare tube which is three inches in diameter. The former is directed downward, either vertical or up to about 45 degrees from the vertical; the latter is directed upward. Thus, a vehicle launched from the garbage tube must be negatively buoyant long enough for the submarine to pass over it completely and then become positively buoyant to rise to the preset hovering depth. This can be accomplished by dropping a weight or purging a flooded chamber at a prescribed time after launching. The vehicle launched from a flare tube is positively buoyant upon launching. Hovering is maintained at a fixed depth by mechanisms similar to those described for vehicles launched from surface ships.

Self-propelled vehicles to simulate submarines are launched from the torpedo tube by swimming out under their own power. These are usually more complex devices and can be programmed to maneuver in azimuth, speed, and depth. In azimuth, a gyro control is used to program course changes throughout the full 360 degrees referred to the course at launching. The depth program is generally confined to a selected stratum, the floor and ceiling of which are controlled by hydrostatic limit switches. Speed changes are effected by a speed control on the propulsion motor. The above program changes must be coordinated in such a manner that they realistically simulate a submarine performing the same maneuver. For example, speed, depth, and course changes must be at a rate reasonable for a submarine. Likewise, turns

must be at a radius commensurate with that of a full-scale submarine and the floor limit should not exceed the diving depth nor should the ceiling limit be such that a submarine would broach or show superstructure.

The self-propelled vehicles are designed to be recovered after exercise use. This is accomplished by running them at negative or neutral buoyancy and dropping a weight at the end of the run, or by running them positively buoyant. In either case, the vehicle rises to the surface when propulsion power is removed. The hovering vehicles are either recoverable or not, depending on economic considerations. The more costly units are recoverable, whereas the less costly ones are designed to flood and sink after the exercise. In the warshot condition, all vehicles are expendable and they are set to sink after a run.

The primary power source in all types of vehicles is a battery. The self-propelled units generally employ a secondary type battery for exercise shots; it is recharged after recovery. For warshots, primary type batteries are used with the electrolyte stored separately; they have indefinite shelf life. The hovering type units employ sea water activated batteries which, if the unit is recovered, are replaced after each use. Since the electrolyte is sea water, there is no problem of maintaining a charged condition and, if stored in a fairly dry atmosphere, the shelf life is nearly indefinite.

Towed vehicles are used primarily from surface ships and can have either electronic or mechanical sound sources. The mechanical sources are of the vibrating bar or motor-driven types. The acoustic output is broadband and provides a higher intensity target than the towing ship. Electronic sources are echo repeaters, broadband maskers, or discrete frequency jammers. The vibrating bar type requires only a mechanical towline. The electronic and motor driven source vehicles require an electrical cable as well and are controlled from the ship. Usually, the latter types are cycled to permit use of the ship sonar without interference from the countermeasure. The vehicles are designed to tow at the proper depth throughout the towing speed range of the ship. Towing from a submarine has been done but is not looked upon favorably. The towing problem is more complex for the submarine since it operates in a volume rather than on a surface. Thus, depth maneuvering, as well as turning, is restricted due to the possibility of the towline becoming fouled with the propellers and the superstructure. However, the newer submarines being more nearly true submersibles, towing from submarines will probably become more common.

23.9 Conclusion

It must be remembered that the useful life of a countermeasure, once put into operation in a war situation, is short at best. Great effort is expended to

overcome any advantage an enemy derives from his countermeasures. Similarly, the advantage of a new measure is limited to the time required to effectively counter it. New measures breed new countermeasures and vice versa. To be sure, both measures and countermeasures may have some residual value in certain circumstances after they have been effectively bypassed by newer techniques. However, space limitations usually legislate against carrying both the old and the new; the old is shunted aside.

Since the measure-countermeasure game is very dynamic and the enemy controls the countermeasure to a large degree by the type of measure he employs, it is generally desirable to design countermeasures to operate against equipment types rather than specific equipments. Even if one could assume detailed knowledge of the susceptibility of an enemy measure (one usually cannot), this would still be true. If countermeasures are designed only against specific equipments, a slight change in the measure could render a countermeasure totally useless. For example, a jammer designed to take full advantage of a given echo-ranging sonar by 'ransmitting the exact frequency and bandwidth accepted by the sonar would be ineffective should the sonar frequency be changed. By the time the jammer could be modified and readied for fleet usage, many ships and lives might be lost. Had the jammer been designed for use against sonars echo ranging within a broad band of frequencies, its utility would not have been lost completely, if at all. Covering a broad frequency band creates problems as far as sensitivity and output power are concerned, but these are compensated for, to a degree, by the one-way path loss as against the two-way loss of the sonar.

Economic considerations generally preclude large reserve stocks of countermeasures. Some are required to meet the immediate needs when a war commences and others for routine training requirements. With limited financial and manpower resources, a balance must be struck between research and development effort, fleet training exercises, and a reserve stockpile. This can be done by maintaining a continuing research and development effort to keep abreast of the latest trends in measures and tactics as they affect future countermeasures development. Simultaneously, the latest countermeasures proven effective in fleet evaluation can be procured for reserve and the latest developments procured for fleet evaluation. When evaluated equipments are placed in production, the reserve stock would be issued for fleet training exercises. Thus, a three phase cycle would be in operation allowing production, evaluation, and development to proceed side-by-side. Fleet usage would provide feedback data upon which improvements could be based. At the same time, the sonar and weapons designers would be better able to come up with equipments less susceptible to countermeasures. Sonar operators would benefit in like manner.

When war starts, or becomes imminent, we would be endowed with a stockpile of ideas and techniques, manpower and physical plant capabilities, a trained fleet, and a limited stockpile of countermeasures.

23.10 Bibliography

This listing is based on the basic function performed by an equipment or the function under investigation. A few reports are listed in more than one functional area since they cover the areas rather completely with no particular emphasis on one over the other. Similarly, there is a fundamental grouping of reports which fit no particular functional area but which are closely related to countermeasures or the field of underwater acoustics.

To include all reported work performed in the field would be an arduous task hardly worthy of the effort. Much work not listed is referenced in the included reports since these are, in general, representative of the significant effort since World War II. Information on work carried out during World War II can be obtained from the Summary Technical Reports of Division 6, National Defense Research Committee, Office of Scientific Research and Development.

23.10.1 Fundamental

"Fundamentals of Sonar," J. W. Horton, NavShips 92719, U.S. Naval Inst., Annapolis, Maryland, 1957 (Unclassified)

"A Summary of Underwater Acoustic Data," R. J. Urick and A. W. Pryce
> Part I—Introduction—July 1953
> Part II—Target Strength—December 1953
> Part III—Recognition Differential—December 1953
> Part IV—Reverberation—February 1954

> Department of the Navy—Office of Naval Research (CONFIDENTIAL)

"High Frequency Acoustic Reflections from Submarine and Submarine Wakes," U.S. Naval Ordnance Laboratory, Report 2601, August 1, 1953 (CONFIDENTIAL)

"Long Range Listening from A Submarine," Scripps Institution of Oceanography, S10 Reference 52-39, August 1, 1952 (SECRET)

"Manual for Estimating Echo Range," Department of the Navy, BuShips, NavShips 900,196, March 16, 1959 (CONFIDENTIAL)

"Submarine Evasion Devices Manual," Commander-in-Chief, U.S. Fleet, Report P-0010, June 21, 1945 (SECRET)

"Analysis of the Effectiveness of the Anti-Submarine Submarine," Chief of Naval Operations—Operations Evaluation Group, Report 573, January 25, 1957 (SECRET)

23.10.2 Detection

"Long Range Passive Acoustic Detection of Torpedoes by Merchant Ships," An Annotated Bibliography, U.S. Navy Mine Defense Laboratory, Technical Paper, TP90, August 1958 (SECRET)

"Sonar Detection of Torpedoes," U.S. Navy Underwater Sound Laboratory, Report 291, August 29, 1955 (CONFIDENTIAL)

"Project PUFFS, Final Report," U.S. Naval Ordnance Laboratory, Report 4242, October 22, 1956 (SECRET)

"Carrier Borne Sonar for Torpedo Evasion," Chief of Naval Operations—Operations Evaluation Group, Report 579, April 24, 1957 (CONFIDENTIAL)

"Narrow-Band Analysis of Low Frequency Signals of Submarines and Surface Ships," Pennsylvania State University, Ordnance Research Laboratory, Report TN 24,350G-9, June 26, 1957 (SECRET)

"An Investigation of the Fine Structure of Torpedo Sounds for Tactical Applications," Final Report on Contract NOnr-1250(00), Melpar, Inc., August 1, 1954 (SECRET)

23.10.3 Interception

"QXB Sonaramic Receiver Instruction Book," NavShips 91076, April 6, 1948

"QXB-2 Sonaramic Receiver Instruction Book," NavShips 92289, Radio Corp. of America, July 30, 1954 (CONFIDENTIAL)

"QXB-3 Sonaramic Receiver Instruction Book," NavShips 92481, Panoramic Radio Corp., April 11, 1959 (CONFIDENTIAL)

The Development of the Underwater Intercept Receiving Set, Q-89," E. L. Hixson, U.S. Navy Electronics Laboratory, Report 533, September 29, 1954 (CONFIDENTIAL)

"AN/WLR-2 Underwater Intercept Receiver," NavShips 93009(A), Instruction Book General Electric Company, (HMED), Contract NObsr 72553, August 23, 1957 (CONFIDENTIAL)

"Evaluation of AN/WLR-2 Underwater Intercept Receiver for Service Use; Information on" Final Report on Project OP/S447/J15, Commander Operational Development Force, November 17, 1958 (CONFIDENTIAL)

"Underwater Intercept Receiver Equipment AN/WLR-2(XN-1)," Final Report (Engineering) General Electric Company, HMED, December 10, 1958 (CONFIDENTIAL)

23.10.4 Masking and/or Jamming

"Sonar Countermeasures Equipment—Summary," V. G. McKenney, K. E. Geren, and W. C. Hubbard, U.S. Navy Electronics Laboratory, Report 472, September 7, 1954 (SECRET)

"Preliminary Results of Sonar Countermeasures Technical Evaluation," P. Hulsveld, Jr., U.S. Navy Electronics Laboratory, Report 751, November 9, 1956 (SECRET)

"Technical Evaluation of Sonar Countermeasures," Technical Film Report Snips 10-56, P. Hulsveld, Jr., U.S. Navy Electronics Laboratory, February 1957 (SECRET)

"Evaluation of Performance of Sonar Countermeasures Systems," P. Hulsveld, Jr., M. D. Papineau, C. C. Routh and W. A. Sauer, U.S. Navy Electronics Laboratory, Report 807, January 30, 1958 (SECRET)

"NAC Beacon Tests," D. J. Evans and V. G. McKenney, University of California, Division of War Research, Report SM-215, May 19, 1944 (CONFIDENTIAL)

"NAC Beacon Tests in 14th Naval District," V. G. McKenney and W. C. Beckley, University of California, Division of War Research, Report SM-240, June 30, 1944 (CONFIDENTIAL)

"NAC Beacon," V. G. McKenney, et al., University of California, Division of War Research, Report S-243, August 15, 1944 (CONFIDENTIAL)

"Final Report on Engineering Evaluation of the NAC-1 Sound Beacon," K. E. Geren and W. A. Sauer, U.S. Navy Electronics Laboratory, Report 293, May 1, 1952 (CONFIDENTIAL)

"NAC-1 Sound Beacon Instruction Book," NavShips 91673, Bell Air Electronics Corp., May 21, 1952 (CONFIDENTIAL)

"Rapid Response Echo Masker," V. G. McKenney and K. E. Geren, U.S. Navy Electronics Laboratory, Report 152, February 16, 1950 (SECRET)

"Evolution of NAH/NAD-10 Mobile Beacon," K. E. Geren, C. C. Routh, U.S. Navy Electronics Laboratory, Report 454, December 14, 1953 (CONFIDENTIAL)

"Final Engineering Evaluation of NAH (Canister type) Sonar Beacon," W. A. Sauer, U.S. Navy Electronics Laboratory, Report 453, December 28, 1953 (CONFIDENTIAL)

"Portable Shipborne NAH Sonar Responder (XG-1)," K. E. Geren and J. W. Ingels, U.S. Navy Electronics Laboratory, Report 467, February 15, 1954 (SECRET)

"Canister NAH Beacon," K. E. Geren and J. W. Ingles, U.S. Navy Electronics Laboratory, Report 538, March 3, 1955 (CONFIDENTIAL)

"AN/SLQ-3 Sonar Beacon Instruction Book," NavShips 92720, Ultrasonic Corp., March 8, 1956 (CONFIDENTIAL)

"A Multiple Sonar Masking/Jamming Device," K. E. Geren, W. A. Sauer, and D. A. Young, U.S. Navy Electronics Laboratory, Report 879, July 1, 1958 (CONFIDENTIAL)

"Expendable Underwater Acoustical Countermeasure, NAE Beacon Mark I, Mod 6," R. P. Crist, David Taylor Model Basin, Report C-605, January 1954 (CONFIDENTIAL)

"Expendable Underwater Acoustical Countermeasure, NAE Beacon, Mark 3, Mod 1, R. P. Crist, David Taylor Model Basin, Report C-725, January 1956 (CONFIDENTIAL)

"Criteria for the Design of Bubble Screening Systems for the Suppression of Underwater Sound," David Taylor Model Basin, Report C-605, February 1957 (CONFIDENTIAL)

"Mechanical Performance, Acoustic Output and Life Tests of HARP Device," U.S. Navy Mine Defense Laboratory Report 77, June 1956 (SECRET)

23.10.5 Deception

"Sonar Countermeasures Equipment—Summary," V. G. McKenney, K. E. Geren, and W. C. Hubbard, U.S. Navy Electronics Laboratory, Report 472, September 7, 1956 (SECRET)

"Preliminary Results of Sonar Countermeasures Technical Evaluation," P. Hulsveld, Jr., U.S. Navy Electronics Laboratory, Report 751, November 9, 1956 (SECRET)

"Technical Evaluation of Sonar Countermeasures—Technical Film Report," Ships 10-56, P. Hulsveld, Jr., U.S. Navy Electronics Laboratory, February 1957 (SECRET)

"Evaluation of Performance of Sonar Countermeasures Systems," F. Hulsveld, Jr., M. D. Papineau, C. C. Routh, and W. A. Sauer, U.S. Navy Electronics Laboratory, Report 807, January 30, 1958 (SECRET)

"Completion Report for NAD-6 Sound Beacon," University of California, Division of War Research, Report U-451, October 1946 (CONFIDENTIAL)

"Preliminary Instruction Manual for NAD-6A Sound Beacon," University of California, Division of War Research, Report SM-332, July 20, 1945 (CONFIDENTIAL)

"Preliminary Instruction Manual for NAD-10 Prime Sound Beacon," University of California, Division of War Research, Report SM-329, July 10, 1945 (CONFIDENTIAL)

"Completion Report for NAD-6 and NAD-10," University of California, Division of War Research, Report File 09.455, October 30, 1945 (CONFIDENTIAL)

"Completion Report on Consulting Services for NAD-10 Sound Beacons," University of California, Division of War Research, Report File M-444, September 28, 1946 (CONFIDENTIAL)

"Completion Report for NAD-10 Sound Beacon," University of California, Division of War Research, Report 452, October 1, 1946 (CONFIDENTIAL)

"Preliminary Instruction Manual for NAD-10A Sound Beacon," University of California, Division of War Research, Report M-412, October 1, 1946 (CONFIDENTIAL)

"Pattern Running Control for Mobile Beacons," W. C. Hubbard, et al, U.S. Navy Electronics Laboratory, Report 294, September 9, 1952 (CONFIDENTIAL)

"Service Test Report, Engineering Evaluation of NAD-10B Mobile Beacon," K. M. Cologne, Jr., U.S. Navy Electronics Laboratory, Report 528, November 22, 1954 (CONFIDENTIAL)

"Final Report on NAD-10B Submarine Simulating Decoy," W. C. Hubbard, U.S. Navy Electronics Laboratory, Report 769, March 1957 (CONFIDENTIAL)

"AN/SQQ-9 Sonar Beacon Instruction Book," NavShips 91863, Hoffman Laboratories, Inc., February 19, 1953 (CONFIDENTIAL)

"Engineering Evaluation of the AN/SQQ-9 Sonar Beacon," W. C. Hubbard, U.S. Navy Electronics Laboratory, Report 537, January 11, 1955 (CONFIDENTIAL)

"Final Report on Development of the 19-inch Sonar Beacon," P. Hulsveld, Jr., M. D. Papineau, and C. C. Routh, U.S. Navy Electronics Laboratory, Report 754, March 1957 (CONFIDENTIAL)

"Final Report on Engineering Evaluation of the AN/SQQ-1 and AN/SQQ-2 Sonar Beacons," K. E. Geren and W. A. Sauer, U.S. Navy Electronics Laboratory, Report 349, February 5, 1953 (CONFIDENTIAL)

"Final Report on Engineering Evaluation of the AN/SQQ-3 and AN/SQQ-4 Sound Beacons," K. E. Geren and W. A. Sauer, U.S. Navy Electronics Laboratory, Report 284, March 18, 1952 (CONFIDENTIAL)

"Development of AN/SQQ-5, -6, -7, and -8, Dopplerized Echo Repeaters," K. E. Geren and W. A. Sauer, U. S. Navy Electronics Laboratory, Report 383, June 14, 1954 (CONFIDENTIAL)

"Service Test Report—Engineering Evaluation of the AN/SQQ-7 and AN/SQQ-8 Sonar Beacons," R. J. Palk and K. H. Cologne, U.S. Navy Electronics Laboratory, Report 536, July 20, 1954 (CONFIDENTIAL)

"AN/SQQ-7 and -8 Sonar Beacons," Instruction Book, NavShips 92496, Stone and Smith, Inc., July 6, 1955 (CONFIDENTIAL)

"Evaluation of TONAR (Towed, narrow-band, decoy for Acoustic Torpedoes)" U.S. Navy Mine Countermeasures Station, Report 50, June 1953 (SECRET)

23.10.6 Torpedoes and Torpedo Countermeasures

"Sixth Bi-monthly Mine Defense Symposium on Torpedo Countermeasures and Oceanographic Studies," U.S. Navy Mine Defense Laboratory, Symposium Vol. II, No. 1, February 15, 1955 (SECRET)

"15th Bi-monthly Symposium, Torpedo Countermeasures," U.S. Navy Mine Defense Laboratory, Bi-monthly Symposium, Vol. IV, No. 1, June 16, 1957 (SECRET)

"19th Mine and Defense Symposium, Torpedo Countermeasures and Anti-Submarine Warfare," February 5 and 6, 1958, U.S. Navy Mine Defense Laboratory, Symposium Vol. V, No. 1, 1958 (SECRET)

"Long Term Research in Torpedo Countermeasures, Task No. 1," U.S. Navy Mine Defense Laboratory, Preliminary Report 28, November 28, 1955 (SECRET)

"Long Term Research in Torpedo Countermeasures, Task 2," Preliminary Report, U.S. Navy Mine Defense Laboratory, January 1958 (CONFIDENTIAL)

"Summary Engineering Report on a Study of Torpedo Countermeasures," Vitro Laboratories, Report KLX-10024, Contract NObs 62024, November 1, 1955 (SECRET)

"Summary Engineering Report on a Study of Torpedo Countermeasures," Vitro Laboratories, Report KLX-10059, Contract NObs 62024, June 15, 1957 (SECRET)

"Countermeasures to Acoustic Torpedoes Used as ASW Weapons," K. E. Geren and W. A. Sauer, U.S. Navy Electronics Laboratory, Report 666, March 29, 1956 (SECRET)

"Convoy Protection Afforded by Stern-Towed Project General Type 2 Streamers," U.S. Navy Mine Defense Laboratory, Report 63, January 1957 (SECRET)

"Theoretical Hit and Acquisition Probabilities for ASW Homing Torpedoes Launched from Fixed Wing Aircraft," Chief of Naval Operations, Operations Evaluation Group, OEG Study 509, October 12, 1953 (SECRET)

"A Study of the Potential Effectiveness of Submarine Launched Passive Homing Acoustic Torpedoes as Anti-Ship Weapons," Bureau of Ordnance—Department of the Navy, Report 170, September 29, 1955 (CONFIDENTIAL)

"Hit Probability of a Wake-Detecting Torpedo," Pennsylvania State University, Ordnance Research Laboratory, Report NOrd 7958-293, October 29, 1954 (SECRET)

"Torpedo Trajectory Analysis," Pennsylvania State University, Ordnance Research Laboratory, Report NOrd 7958-294, December 29, 1954 (CONFIDENTIAL)

"Summary of Underwater Acoustic Reverberations and Some Effects on the Performance of Active Acoustic Torpedoes," U.S. Naval Ordnance Test Station, Report 3406, January 4, 1955 (CONFIDENTIAL)

"Sonar—Signal Interference with Acoustic Torpedo Homing," U.S. Naval Ordnance Test Station, Report 4952, September 27, 1955 (CONFIDENTIAL)

"An Auto-correlation Torpedo Guidance System Capable of Homing on Either Radiated Noise or Echo-Ranging Signals," P. Huisveld, Jr., M. D. Papineau, and C. C. Routh, U.S. Navy Electronics Laboratory, Report 913, July 9, 1959 (CONFIDENTIAL)

"Torpedo Mark 32, Mod 2, Surface—Ship—Launched Shots—to—Hit," Commander Operational Development Force, December 2, 1953 (CONFIDENTIAL)

"Project CYCLONE—Acoustic Homing Study," Reeves Instrument Corp., Report on Contract NOas 53-383-c, April 1954 (SECRET)

23.10.7 Lists Compiled Elsewhere

"A Bibliography of Ambient Noise and Underwater Sound Output of Surface Ships and Submarines (revised)," U.S. Navy Underwater Sound Laboratory, Report 235, June 16, 1954 (CONFIDENTIAL)

"Underwater Sound—A Bibliography of Contemporary Literature Available," Office of Naval Research, Surveys of Naval Science, ONR Survey ACR-20, 1955–1956

"Underwater Acoustic Bibliography, U.S. Navy Mine Defense Laboratory, September 1956 (SECRET)

"Underwater Sound, An Annotated Bibliography, Vol. I," Library of Congress, Information Division, August 1956.

"Underwater Sound, An Annotated Bibliography, Vol. II," Library of Congress, Information Division, September 1957

"Abstracts of UCDWR and NEL Reports on Sonar Countermeasures," W. A. Sauer, U.S. Navy Electronics Laboratory, Report 940, November 5, 1959 (SECRET)

Acknowledgements

Part IV

Components

24

Circuits

W. A. EDSON, J. M. PETTIT

Section I: Fundamentals

24.1 Linear Video Amplifiers: Fundamentals

The first portion of this chapter is concerned with linear amplification of small-signal voltages. An ideal amplifier, when supplied with an input signal $e(t)$, will deliver an output voltage $Ae(t)$. The factor A is a constant, and represents the "gain". The waveform of the output voltage is ideally identical in shape to, but larger in amplitude than the input waveform.

The field of application is widespread, including amplifiers for cathode-ray oscilloscopes, television and radar systems, etc. Such amplifiers are commonly called video or wideband amplifiers, and are discussed in most of the standard textbooks (for example, Reference 1), although from the steady-state, sine-wave point of view. In contrast, we shall be concerned with the time response, or transient response, i.e., we apply a square wave to the input of the amplifier and examine the waveform that results at the output. For a supplementary reference that is written from this point of view, and from which many of the examples in the following pages have been taken, see Reference 2. It is, of course, true that the transient and steady-stage responses are interrelated, but the relationship is indirect.

We shall use pentodes,* and in the linear condition. That is to say, we shall have to restrict the tube currents and voltages to a small region of the tube characteristics such that the i_b vs. e_b curves can be assumed parallel

---

*Transistors are also used in linear amplifiers, of course. The general philosophy of circuit design is similar, although the details are different in various respects. For a more complete treatment of both tube and transistor circuits, see Reference 2a.

straight lines, and equally spaced with respect to increments of grid voltage e_g. Then the tube can be replaced by the equivalent circuit shown in Figure 24-1.
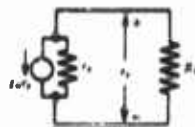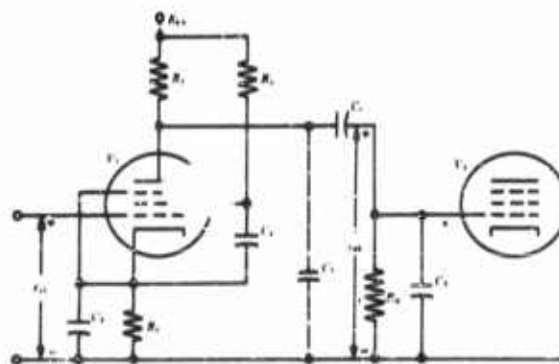
FIGURE 24-1   FIGURE 24-2

Of course, to connect one stage to the next there will be required some extra elements as in Figure 24-2, notably a coupling capacitor C_c to isolate the d-c plate voltage of the first stage from the d-c grid voltage of the next stage, a grid resistor R_g for providing d-c bias connection, a cathode bias combination of R_k and C_k, and finally the stray capacitance of the tube and wiring, C_1 and C_2. This entire circuit is a bit formidable to analyze in formal detail, so we shall take steps to simplify it, much as one does with audio amplifiers deriving "high-frequency" and "low-frequency" equivalent circuits. We shall do the same things on a transient basis.

In passing, take note that all the elements in Figure 24-2 with the exception of R_L are parasitic, i.e., they impair rather than aid in the production of constant gain or voltage amplification. The basic gain, if the effects of the parasitic elements were negligible, can be seen from Figure 24-1 to be:

$$\text{Gain} \triangleq \frac{e_{g_2}}{e_{g_1}} = \frac{-g_m e_g \frac{r_p R_L}{r_p + R_L}}{e_g} \approx -g_m R_L$$

If $r_p \gg R_L$, where $\triangleq$ means "is defined as."

To keep things a bit simpler at the outset, let us omit the cathode bias circuit as a consideration (we shall deal with it later). When the tube is replaced by its equivalent circuit as in Figure 24-1, the network becomes that of Figure 24-3.
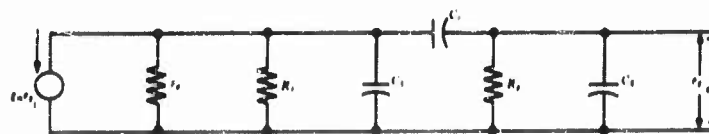
FIGURE 24-3

We are going to test this network with a step function of voltage as the signal at the grid of V_1 (Figure 24-2). We shall find that the circuit of Figure 24-3 can be resolved into two separate circuits, one that completely describes the behavior of the output voltage for small values of time, i.e., during the abrupt rise of the step voltage, and a second one that is sufficient for larger values of time, i.e., during the long interval when the input step voltage is constant at its maximum value. The adequacy of the two circuits is better for "good" amplifiers, i.e., amplifiers that come as close as possible to reproducing the input step at the output terminals. It will turn out that a good amplifier has R_g and r_p much larger than R_L, and C_c much larger than C_1 and C_2.

First consider the abrupt rise of the input step of voltage e_{g_1}. If C_1 and C_2 were absent, there would be nothing to impair the instantaneous rise in the output voltage e_{g_2}. The capacitor C_c cannot change its voltage instantaneously, but it need not do so; if its initial voltage is zero then it acts like a short circuit for instantaneous circuit changes. In fact, C_c is usually so large that there is virtually no change in its voltage during the small but finite time required for the output voltage to rise. The reason that a finite rise time is actually required is that C_1 and C_2 must be charged from zero to the peak value of the step. They are small compared to C_c, so we can draw an equivalent circuit showing them but omitting C_c, or rather, replacing C_c by a short circuit as in Figure 24-4.
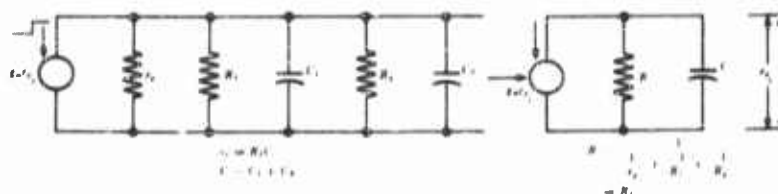
FIGURE 24-4

When the input voltage e_{g_1} is a step function, the output voltage will "rise" exponentially with a time constant RC, as depicted in Figure 24-5.

Now we take the next step. After the output voltage has risen in the comparatively short time required for it to reach its final value $g_m R_L E_o$, the
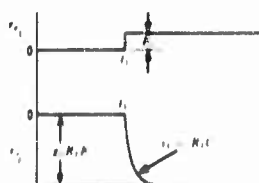
FIGURE 24-5

voltage on capacitor C_o will begin slowly to change. We can see from Figure 24-3 that the output voltage will ultimately decay toward zero. Capacitor C_o will then be charged to the full voltage $g_m R_L E_o$, and capacitor C_y will be discharged to zero. We say this: Because capacitor C_o is so much larger than C_y (and C_1 as well), the current required to discharge C_y is negligible compared to that required to charge C_o, and we shall not bother to show C_1 or C_y in the equivalent circuit governing the charging of C_o. See Figure 24-6, and the resulting waveforms in Figure 24-7.
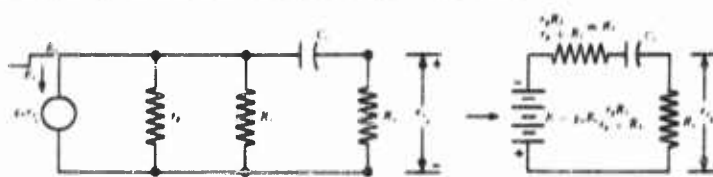
FIGURE 24-6
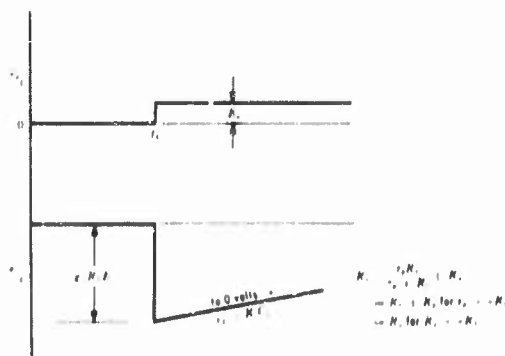
FIGURE 24-7

The best response to the step is provided by the smallest possible $\tau_1 = RC$ and the largest possible $\tau_2 = R_o C_o$. Since the total tube capacitance C is usually predetermined, one can only decrease R in order to make τ_1 smaller; this means decreasing R_L, since R_y must be kept large to preserve a high value of τ_2. But decreasing R_L costs in gain, since the gain at the maximum of the output waveform is $g_m R_L$.

In a similar manner there is a limit to the amount of improvement of the long-time behavior that can be achieved by increasing τ_2. The grid resistor

cannot be made greater than about a megohm, from considerations of grid current, the capacitor C_c cannot be made arbitrarily large, because its physical size will add to the shunting capacitance C, and thus impair the short-time response.

The two circuits of Figures 24-4 and 24-6 correspond to the high-frequency and low-frequency equivalent circuits, respectively, that one derives from steady-state analysis of the elementary resistance coupled amplifier. There is a simple interrelationship for this circuit between the time constant τ_1 that determines the short-time behavior and the frequency f_1 at which the steady-state response has fallen to 70.7 percent of its midband value; this frequency is that for which

$$X_c = R$$
$$= \frac{1}{2\pi f_1 C}$$
$$f_1 = \frac{1}{2\pi} \cdot \frac{1}{RC} = \frac{1}{2\pi} \cdot \frac{1}{\tau_1} \cong \frac{1}{6\tau_1}$$

As an example, if τ_1 is one microsecond; i.e., the step response would rise to 63.2 percent of its final value in one microsecond, the corresponding high-frequency 70.7 percent frequency f_1 would be $\frac{1}{6}$ of one megacycle, or 160 kc.

Similarly, the low-frequency 70.7 percent frequency f_2 is related to the time constant τ_2 that determines the long-time behavior in the transient response as follows:

$$X_c = R_a$$
$$= \frac{1}{2\pi f_2 C_c}$$
$$f_2 = \frac{1}{2\pi R_a C_c} = \frac{1}{2\pi \tau_2} \cong \frac{1}{6\tau_2}$$

As a numerical example involving τ_2 and f_2, suppose that the step-response requirement is that the top of the step "sag" by only one percent in 100 microseconds. Then considering the initial portion of the exponential curve of Figure 24-7 to be a straight line,

$$\frac{1\%}{100\%} = \frac{100 \ \mu sec}{\tau_2}$$

or

$$\tau_2 = 10^4 \ \mu sec = 10^{-2} \ sec$$
$$f_2 = \frac{1}{2\pi \ \tau_2} = \frac{100}{2\pi} = 16 \ cps$$

In the general case of amplifier networks that are more complicated than the simple resistance-coupled case, and indeed in the case of a cascade of several resistance-coupled stages, there is not such a simple relationship between the transient and steady-state responses of an amplifier. Nonetheless, it is true in a rough way that a fast rise in the transient response calls for a high-frequency limit in the steady-state characteristic, and that small sag implies a rather low 70.7 percent cutoff frequency. More specific comments will be made later.
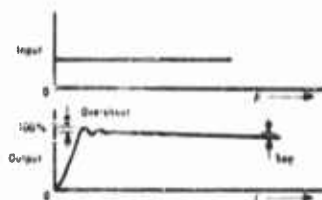
FIGURE 24-8

In general the response of an amplifier to a step of voltage at the input is not the simple combination of the two exponentials in Figure 24-5 and 24-7, but may be more like that in Figure 24-8.

The distortion introduced by the amplifier; i.e., its failure to reproduce the input waveform is seen to comprise several aspects. First, of course, the rise of voltage at $t = 0$ is not instantaneous, but rather there is a finite rise time; more will be said about this below. Secondly, the output waveform does not rise uniformly (monotonically) to its 100 percent level, but instead there is an oscillation superimposed which produces an *overshoot*, expressible in percent. Finally, the output waveform fails to maintain the 100 percent level; there is a sag, which can be expressed as a slope; i.e., percent per second, volts per microsecond, etc., or as the percent sag in a given time interval of particular interest.

The rise time can be defined in several ways, all of them giving about the same numerical results but each being more convenient than the others for some purpose. It is necessary first to distinguish a delay time, which can be regarded separately from rise time. In general, delay time is not considered a distortion since, if all waveforms are preserved but delayed by an absolute time interval, most electronic systems will function properly. The output waveform of Figure 24-8 is reproduced in Figure 24-9, and for comparison there is shown a delayed step with which the actual output can be compared. The delay time is arbitrarily defined as the time for the actual output to reach 50 percent of its basic level.

24.2 Linear Video Amplifiers: Speed of Step Response (Rise Time)

Coming now to the detailed analysis of that portion of the amplifier circuit which determines the short-time response to a step of voltage at the input, we shall be concerned with several aspects:
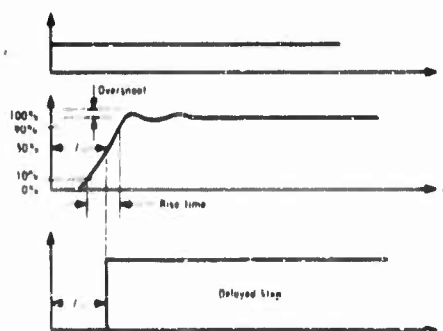
FIGURE 24-9

1) The choice of tube
2) The choice of circuit
3) The behavior of a cascade of stages
4) Relations between step and steady-state response
5) Additive amplifiers
    a) Distributed amplifier
    b) Band splitting amplifier

Notice again that the equivalent circuit which governs the simplest possible amplifier stage, the resistance-coupled network of Figure 24-2, is merely that of Figure 24-4.

If e_{g_1} is a step of voltage, applied at $t = 0$ and with amplitude equal to E_0, the output voltage e_{g_2} is:

$$e_{g_2} = -g_m E_0 R \left(1 - \epsilon^{t/\tau_1}\right) \text{ where } \tau_1 = RC \qquad (24\text{-}1)$$

The magnitude A of the amplification or gain is defined for the maximum (final) value of the output voltage:

$$A = \left| \frac{e_{g_2}}{e_{g_1}} \right|_{t \to \infty} = g_m R \qquad (24\text{-}2)$$

The rise time for the circuit according to the 10 to 90 percent definition for the exponential function of Eq (24-1) is:

$$T_{R} = 2.2\tau_1 = 2.2 \, RC \qquad (24\text{-}3)$$

24.2.1 Choice of Tube

If one sets out to design the resistance-coupled amplifier stage to give both

high gain and a short rise time he is confronted with a contradiction. From Eq (24-2) and (24-3) it is seen that for a given tube, i.e., a given g_m and C, one can increase R to raise the gain but in doing so one lengthens the rise time. This proportionality of gain and rise time can be expressed as a quotient, whose magnitude is constant and depends primarily upon the tube:

$$\frac{\text{Gain}}{\text{Rise Time}} = A/T_H = g_m/2.2C \qquad (24\text{-}4)$$

The capacitance C ($= C_i + C_a$ in Figure 24-2) includes both input capacitance C_i and output capacitance C_a of the tube (assuming both tubes associated with the interstage network are of the same type), together with the stray wiring capacitance. The latter can usually be made small compared to the tube capacitance, and in any case it is apparent that, if two tubes have equal g_m but different C, the one with the smaller C will be better. Table 24-I shows listings of the transconductances (g_m), the input capaci-

TABLE 24-I

Tube	C_i ($\mu\mu f$)	C_a ($\mu\mu f$)	C ($\mu\mu f$)	g_m μmho	Gain/Rise Time (per μsec)
6AK5	4.0	2.8	10.8	5000	210
6AG5	6.5	1.8	12.3	5000	185
6AH6	10.0	2.0	16.0	9000	256
6AU6	5.5	5.0	14.5	5200	163

tance C (which includes 4 $\mu\mu f$ for stray wiring capacitance), and the A/T_H quotient, which becomes a sort of figure-of-merit *for the tube* in a resistance-coupled amplifier circuit. A graphical tabulation of many tubes is shown in Figure 24-10.

Although Eq (24-4) and Table 24-I are derived from the properties of the elementary resistance-coupled amplifier stage, it will turn out that with more complicated networks the same figure of merit will apply. Better circuits will give more gain for the same rise time, but $g_m/2.2C$ is still a common multiplier for all.

24.2.2 Choice of Circuit

It might well be expected that by going to a more sophisticated circuit than the elementary resistance-coupled interstage network one could achieve better performance in terms of more gain for a given rise time, i.e., a higher gain/rise-time quotient. The g_m/C ratio of the tube is a universal factor that appears in the gain/rise-time quotient for all the circuits, so we should eliminate this as a factor in comparing alternative circuits to be used with
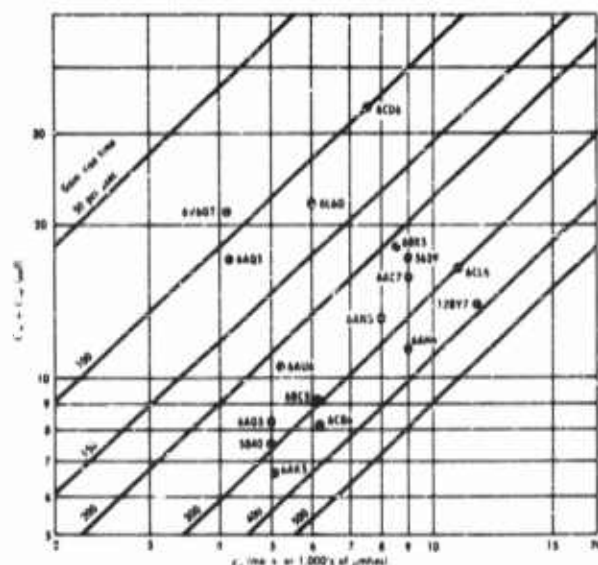
FIGURE 24-10

the same tube. Since the resistance-coupled circuit is the simplest, we can use it as a reference and divide the gain/rise-time quotient for any other circuit by $g_m/2.2C$. This will give then a *relative speed* or figure of merit *for the circuit*. A more complicated figure of merit which puts in a factor for the overshoot introduced by many circuits is given by Palmer and Mautner. To use this figure it is necessary to know the acceptable limit of overshoot for the service intended, e.g. perhaps 2% for television, as suggested by the authors (Reference 3). There are several basic network structures that can be used; the more complicated ones give greater speed, but the designer must decide where to draw the line, for beyond some point the added complexity does not give enough improvement to justify the added difficulty of design and adjustment (particularly in a manufacturing or field servicing situation where unskilled workers are involved.)

Shunt-Peaked Circuit

The first step in circuit refinement beyond the elementary resistance-coupled circuit leads one to the so-called shunt-peaked circuit*, shown in

*The name is derived from steady-state considerations, where the parallel resonance of L and C tends to produce a peak in the curve of amplification versus frequency. By way of contrast, the *series*-peaked circuit, which enjoyed popularity for a time, employes an inductance in series with the coupling capacitor C_p.
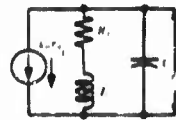
Figure 24-11. This circuit provides a substantial increase in speed relative to the resistance-coupled circuit, and with little increase in complexity. It is probably the most widely used of the circuits discussed here.

It is necessary to specify a parameter that defines the value of L relative to R_L and C. Notice that L is a variable to be adjusted after R_L has been chosen to provide the desired gain, since the final value of gain for the circuit is $g_m R_L$ as before, and C is fixed by the tube. Let this factor be called m, after Valley and Wallman (Reference 2, page 73),* defined by the relationship:

$$m = L/R_L^2 C \qquad\qquad (24\text{-}5)$$

As the factor m is increased from zero (corresponding to the simple resistance-coupled case) to a value of 0.6 by increasing L for a given combination of R_L and C, the step response curves that result are as shown in Figure
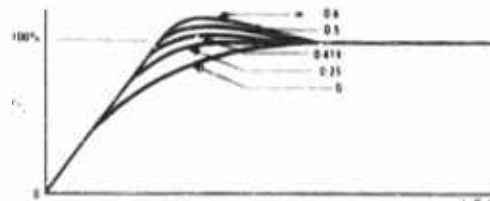
FIGURE 24-12

24-12 It can be seen that as m is increased the rise time decreases, with overshoot appearing for values of m greater than 0.25. A physical interpretation of the effect of adding the inductance is that in order to charge the capacitor C as rapidly as possible, the maximum current should flow from the generator into C. Without L, the initial current does flow entirely in C, but as the voltage builds up, more of the generator current is by-passed into the resistor R_L. Adding L slows up the increasing current in the R_L branch. Eventually the current buildup in R_L is slowed so much that the capacitor voltage overshoots its final value (note that if L were infinite the R_L branch would be an open circuit, and the capacitor voltage could increase indefinitely at a rate $I/C = g_m E_0/C$).

The figure of merit of the shunt-peaked circuit, i.e. its speed (10 to 90 percent rise) relative to the resistance-coupled circuit increases with m as

*Other writers use the same or similar factors Terman (Reference 1) calls it Q_y, since it is the Q of the circuit at a frequency of f_y where $X_c = R_L$ (we have called this f_1; see Section 24-1). Also $m = (Q_n)^2$, where Q_n is the circuit Q at the resonant frequency $f_n = 1/(2\pi\sqrt{LC})$
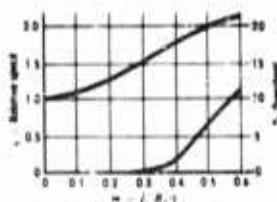
FIGURE 24-13

illustrated in Figure 24-13. Also shown is the curve of overshoot, which is zero up to $m = 0.25$ and then increases with m.

From the figure it can be seen that the most beneficial range of m is from 0 to 0.25, where the relative speed increases from 1.0 to 1.4 without any overshoot appearing. Beyond an m of 0.25 the overshoot commences, with both η and the overshoot increasing but the latter more rapidly.

Discussions based upon steady-state analysis sometimes mention a value of $m = 0.414$ as critical peaking, or critical compensation. It turns out that this value of m is the crossover point between a frequency response curve that falls off uniformly at the high-frequency end—as does the resistance-coupled case—and one that has peak or hump. But from the standpoint of the transient response to a step, this value of m has no significance. (Figure 24-12). Similarly, $m = 0.322$ can be shown to give—in the steady-state response—an optimum linearity of phase shift versus frequency, but again not unique in the transient response.

It is of interest to point out at this juncture that certain proposals have appeared in the literature for two- and three-stage amplifiers based on staggered (nonidentical) values of Q for the stages (Reference 4), or on feedback for two stages (Reference 5). These arrangements will in general introduce substantial overshoot.

Two-Terminal Linear-Phase Network

A network credited to S. Doba of the Bell Telephone Laboratories is the two-terminal linear-phase network of Figure 24-14, so called because of its linear relationship between steady-state phase shift and frequency. With the element values as given, the network has a step response that is 1.77 times faster than the resistance-coupled counterpart, i.e., the rise time is $2.2R_LC/1.77$. There is a small overshoot of approximately 1 percent. Notice that there may be coil capacitance in the shunt-peaked circuit, giving the equivalent of Figure 24-14. Adjusting $C_a = 0.22C$ gives an optimum performance.
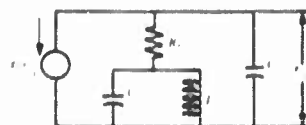
FIGURE 24-14
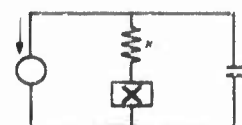
FIGURE 24-15

FIGURE 24-16

Other Two-Terminal Networks

Further complexity of structure can be added to the two-terminal form of network by considering the reactance X in Figure 24-15 to be an arbitrarily extensive arrangement of L and C as in Figure 24-16. This situation has been explored by Elmore (Reference 6), with results as follows: (a) the improvement in speed of rise as each new element (L or C) is added diminishes rapidly after the first one or two; (b) an ultimate improvement factor of 2.12 for an infinite number of elements is suggested by the analysis, though not proved conclusively.

Four-Terminal Networks

A whole family of four-terminal networks can be devised which give substantial improvement over the two-terminal variety. Added complexity results, however, and the response of the networks to a step is sensitive to the ratio of C_1/C_2.

The typical examples are illustrated in Figures 24-17 and 24-18. The former is called the "four-terminal linear-phase network", and provides a relative speed of 2.48 over the resistance-coupled counterpart. A $\frac{1}{4}$ ratio of C_1/C_2 is assumed in the design.
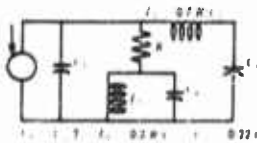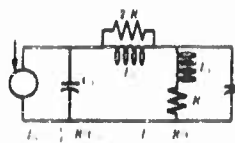
FIGURE 24-17   FIGURE 24-18   FIGURE 24-19
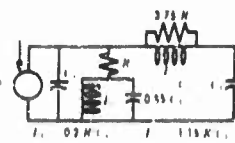
The circuit of Figure 24-18 is commonly called the "series-shunt-peaked network". It assumes a 1:1 capacitance ratio, but is less sensitive to deviation from this value.
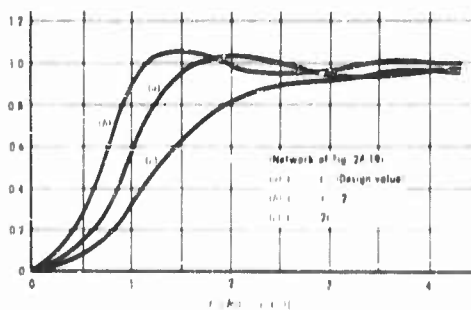
FIGURE 24-20

In Figure 24-19 is a third circuit, this one a 1/1 ratio of capacitance. Failure to realize this ratio results in response waveforms as shown in Figure 24-20.

The relative speed of the four-terminal networks is substantially greater than that of the two-terminal forms. The circuit of Figure 24-17 provides a speed of increase of $\eta = 2.48$, while those of Figures 24-18 and 24-19 give values of 2.06 and 2.10, respectively. The maximum speed of a four-terminal network has not been conclusively established, (Reference 2, pp. 81, 82, and Reference 7) but is probably in the neighborhood of four. As in the case of two-terminal circuits, the actual networks in practical use fall short of the maximum, but provide much improvement over the resistance-coupled form. More complicated structures than those shown here can be devised, but the added speed comes slowly with the extra elements required. A class of interstage networks based upon filter theory was presented in a classic paper by Wheeler, in Reference 8.

Amplifier Stages in Cascade

In general, a single stage of amplification is not adequate to meet the gain requirements of a system, and hence several stages will be connected in cascade. The question arises as to the over-all response of the system when the responses of the individual stages are known. Of course, the entire system could be analyzed as a formal circuit problem, but there are some general rules of great value.

The nature of the transient response of several stages (identical) in cascade is illustrated in Figures 24-21 and 24-22. The first is from Valley and Waliman, and the other from Bedford and Fredendall (Reference 9).
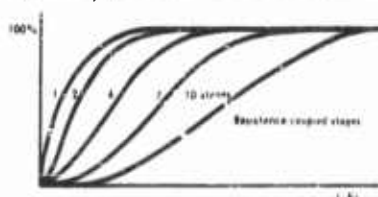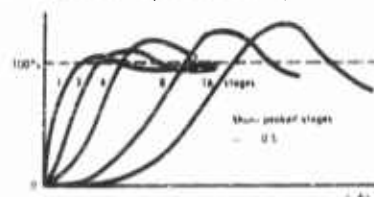
FIGURE 24-21   FIGURE 24-22

The two figures show the general properties of increasing delay time and rise time as the number of stages is increased. Moreover, if there is overshoot in a single stage, the amount of this increases as stages are added. On the other hand, if there is no overshoot in a single stage, the adding of stages will not introduce overshoot (Figure 24-21).

The quantitative details are of interest. Various attempts have been made to analyze the problem, some based upon an empirical study of cas-

cades of particular circuits.* The analysis is quite laborious, even if the
network functions and their inverse transforms are simple; this is because of
the arbitrary definition of the rise time in terms of the 10 and 90 percent
levels. For instance, the time function for a cascade of resistance-coupled
stages, whose response in time is given in Figure 24-21, is simply:

$$e_n(t) = 1 - \epsilon^{-t} \sum_{r=0}^{n-1} \frac{t^r}{r!} \qquad (24\text{-}6)$$

$$(\text{Note: } 0! \overset{\Delta}{=} 1)$$

Equation (24-6) describes the family of curves in Figure 24-21, where t in
the equation is the normalized variable t/RC in the figure. Unfortunately
there is no correspondingly simple expression to describe the rise time T_R
as a function of the number of stages; one must compute the function $e_n(t)$
for each value of n, determining the time between the 10 and 90 percent
levels. Valley and Wallman give the results for values of n up to 10. These
are given here in Table 24-II.

TABLE 24-II

Number of Stages, n	1	2	3	4	5	6	7	8	9	10
Rise Time T_R divided by 2.2 RC	1.0	1.5	1.9	2.2	2.5	2.8	2.9	3.1	3.3	3.5

The results of empirical studies such as these can be summarized into
several working rules, as formulated by Valley and Wallman:

1. In circuits having little or no overshoot, the over-all rise time T_{R_n} is
given by (Reference 6):

$$T_{R_n} = \sqrt{T_{R_1}^2 + T_{R_2}^2 + T_{R_3}^2 + \ldots + T_{R_n}^2} \qquad (24\text{-}7)$$

1a. When the stages are identical, each having a rise time T_{R_1}, the over-
all rise time is:

$$T_{R_n} = \sqrt{n}\, T_{R_1} \qquad (24\text{-}7a)$$

2. In circuits having little or no overshoot, the total overshoot for n
stages is essentially that of a single stage.

*Reference 2, pp. 65-66, 77-78 and References 9, 10, and 11.

3. If the overshoot of a single stage is in the order of 5 or 10 percent, then the total overshoot goes as the square root of the number of stages.

3a. When the stage overshoot is 5 or 10 percent, the total rise time is somewhat less than that given by Rule 1.

4. An interesting result (References 6, 12, and 13) can be obtained by combining rule 1a with the following expression for the over-all gain A_t in terms of the stage gain A_i.

$$A_n = (A_t)^n \tag{24-8}$$

For a specified value of gain A_n there is a minimum value of rise time T_{u_n} which obtains, regardless of tube type, when:

$$n = 2 \ln A_n \tag{24-9}$$

$$A_t = \sqrt{e} = \sqrt{2.72} = 1.65 \tag{24-10}$$

For a particular tube, the value of this minimum T_{u_n} is:

$$\min T_{u_n} = \frac{\sqrt{2e \ln A_n}}{\eta g_m / 2.2C} \tag{24-11}$$

The relationships given above have some practical limitations, and require some judicious interpretation and application. First note that n has the same value, regardless of tube type or circuit type; for instance, if A_n is 10^5 (100db), Eq (24-9) says that 23 stages are called for, whether one uses 6AU6 or 6AH6 tubes. But it does *not* say that the same minimum rise time results in either case; from Table I and Eq (24-11) it will be found that the 6AU6 would give 256/163 or 1.57 times as great a rise time as the 6AH6.

Also, the minimum of the rise-time function is a broad one, and one can violate the minimum condition by quite a margin without serious detriment to the over-all rise time. Elmore provides an example of a 6AC7 amplifier ($g_m = 0.009$, $C = 22 \mu\mu f$, $\eta = 1.5$), in which the 23 stages required for the 100 db gain give a rise time of 0.032 μsec.; yet with only nine stages to give the same gain the rise time is only 0.044 μsec. Thus, the rise time is impaired only 37.5 percent for a 48 percent reduction in the number of stages. Moreover, the larger load resistor in the latter case (400 ohms instead of 180 ohms) permits a larger output voltage to be realized from the amplifier (222 percent greater). This brings us to a new topic.

Output Stages

Most amplifiers have, in addition to the requirement for a certain amount

of gain between input and output terminals, a requirement on the amount
of voltage (or power) that may be needed at the output. The maximum
output voltage is limited by the amount of plate current that can flow
through the load resistor of the last stage. The highest value that can be
achieved—without regard to the requirements of linearity—is that of a step
function that carries the plate current from zero to the maximum rated
value for the tube. This rated value differs from tube to tube, and one would
tend to choose a tube with a high current rating. But, since the rise time of
the output stage enters into the total rise time of the amplifier, one must
consider this factor as well. Several tubes can therefore be compared on the
basis of a new figure of merit suitable for output stages as proposed by
Valley and Wallman (Reference 2, pages 103, 104), this is the ratio of
maximum voltage output to rise time, and depends upon the output capaci-
tance C_o of the tube and the capacitance of the load C_L. As an example,
several tubes are compared in Table 24-III for a 20 $\mu\mu f$ load, such as might
be encountered with the deflection plates of a cathode-ray tube.

$$E_{max} = I_{b_{max}} R_L \tag{24-12}$$

$$T_R = 2.2 R_L (C_o + C_L) \tag{24-13}$$

$$E_{max}/T_R = I_{b_{max}}/2.2(C_o + C_L) \tag{24-14}$$

While the data of Table 24-III give a relative rating to the tubes listed, the
actual number of volts per microsecond that is obtainable may vary with the
practical circumstances. The values listed assume either that the tube carries
rated current with no signal, and that a negative-going step cuts off the
current entirely, or the opposite of this, i.e., the current is cut off in the
quiescent state but turned full-on by a positive-going step at the grid. If the
amplifier must accommodate steps of either polarity, the quiescent operating
point would have to be chosen as approximately $I_{b_{max}}/2$.

TABLE 24-III

Tube	$I_{b_{max}}$ (ma)	C_o ($\mu\mu f$)	E_{max}/TR v/μsec
6AU6	10	5	182
6AK5	10	3	200
6CL6	30	5.5	248
6AQ5	45	8.2	725
6V6GT	45	7.5	750
6L6G	75	10	1130

A special case exists in which the signals are known to be always pulses of a duration that is short compared to the interval between pulses, i.e. pulses of low "duty cycle". In such a case it may be possible to exceed $I_{b_{max}}$ on the positive peaks if the tube rating is based upon heating, i.e. plate dissipation. Sometimes the rating is based on emission limitations or grid current. Each tube needs to be treated as a separate problem.

Occasionally an output stage must operate into a low-resistance load, such as a coaxial cable used to transmit the signal to a distant location. Such a cable is usually terminated in its characteristic resistance R_0, and so the impedance seen from the amplifier is simply a resistance of this value. From consideration of the long-time (or low-frequency) response—to be taken up in Section 24.3—an unreasonably large coupling capacitor would be required with a low value of R_g (notice that R_0 corresponds to R_g in the analysis associated with Figure 24-2. Hence, the output stage is usually operated either
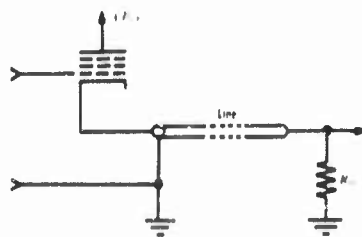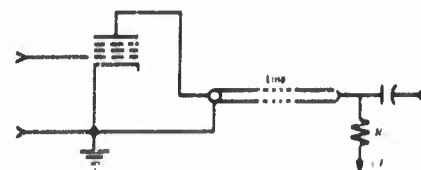
FIGURE 24-23   FIGURE 24-24
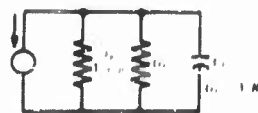
FIGURE 24-25

as in Figure 24-23 or 24-24. The cathode-follower arrangement of Figure 24-23 has the advantage that the coaxial line is at a low d-c potential. Otherwise the two circuits are comparable so far as transient response is concerned. The equivalent circuit of the cathode follower is shown in Figure 24-25. For an interesting commentary, see Reference 14.

If the output stage must drive a capacitive load, such as an oscilloscope, for instance, the cathode follower will provide a smaller rise time (at the expense of small gain, however, but we are not considering gain for the output stage. We would if we have alternative ways of providing the needed output voltage.) In contrast with Eq (24-13), the rise time for the cathode follower is:

$$T_R = 2.2 R_L C_L / (1 + g_m R_L)$$
$$= 2.2 C_L / (G_L + (1 + \mu) r_p) \qquad (24-15)$$

It should be pointed out that with a large capacitive load, it is sometimes

feasible to use several output tubes in parallel with improved performance. The current is doubled by adding the second tube, but the total capacitance increased by a smaller percentage.

Also it may be that for given specificat. a of output voltage and rise time there will be several tube possibilities that would be satisfactory. In such a case it would be reasonable to introduce other factors into the comparison such as gain and input capacitance, the latter influencing the rise time and gain of the preceeding stage.

Transient versus Steady-State Response

Here we have a topic of long standing theoretical interest, and one of considerable practical importance. The state of our knowledge is substantial, but unfortunately not reducible to a few simple axioms. Although a full-scale recounting of the published papers is not practicable here, the results can be summarized and a few common misconceptions pointed out.

a. *Rise Time versus Bandwidth.* For many years the term wideband has been used to describe the type of amplifier one builds in order to obtain a fast response to a step transient, especially in the television art. As was pointed out (Section 24.1) for the simple resistance-coupled circuit there is a correspondence between the rise time and the high-frequency limit of the amplifier at which the steady-state amplitude response is down to 70.7 percent of the "midband" response. The various empirical studies have shown a general relationship to exist as follows:

$$T_R B = 0.35 \text{ to } 0.45 \qquad (24\text{-}16)$$

where T_R = rise time, 10–90 percent
        B = bandwidth, from 0 to upper 3 db frequency

In Eq (24-16) the value of 0.35 matches best those circuits where the overshoot is small or zero, while 0.45 corresponds to overshoots of, say, 5 percent or greater.

Theoretical analysis of two idealized situations yields values of $T_R B$ that compare favorably with Eq (24-16). The first of these is the so-called ideal filter, having a characteristic as shown in Figure 24-26. Associated with the amplitude characteristic as shown is a linear phase; i.e., constant time delay for all frequencies transmitted. Such a response is not physically realizable; this is proved by Valley and Wallman (Reference 2, pp. 721-723), but is also apparent from the fact that when a step function is applied at $t = 0$ the computed response shows finite output prior to this time. The nature of

FIGURE 24-26
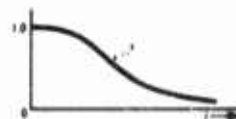
FIGURE 24-27

the step response is a sine-integral function, yielding always 9 percent over-shoot and

$$T_\kappa B = 0.51 \qquad (24\text{-}17)$$

Another amplitude response of theoretical interest is the so-called gaussian function shown in Figure 24-27. Once again we associate a linear phase characteristic with this amplitude response, and once again the combination is not physically realizable (although the amplitude characteristic is achieved in the limit by an infinite number of resistance-coupled stages). The response of a system of this kind to a step function is also a gaussian function, possessing zero overshoot, and:

$$T_H B = 0.41 \qquad (24\text{-}18)$$

The conclusions to be drawn from these results are that for a given kind of circuit, and for the same amount of overshoot, a faster rise is obtained with a greater bandwidth. However, merely increasing the bandwidth without regard to overshoot does not necessarily lead to "better" response. Thus, taking a given amplifier, whose amplitude response may be a gradually decreasing function of frequency and attempting to speed it up by adding compensating elements in order to make its response approach that of Figure 24-26 will indeed speed up the amplifier because of the higher value of the $T_H B$ product, but the resulting overshoot may render the amplifier worthless for the intended application.

b. *Transient Distortion versus Steady-state Distortion.* An ideal amplifier from the transient view would have zero rise time and no overshoot. An ideal amplifier from the steady-state view would have an amplitude response that would be constant to infinite frequency and phase shift proportional to frequency. Failure to achieve these ideals is termed distortion. The transient distortion is described in terms such as rise time and overshoot, whereas the steady-stage distortion can be described as amplitude distortion* and phase

*Sometimes called frequency distortion by authors who use amplitude distortion for nonlinear effects.

distortion. A distortion in either the amplitude or phase characteristics will lead to transient distortion. This fact is not obvious and several analyses will be found in the literature that examine the relative importance of the types of distortion; we shall shortly consider two of these analyses.

Actually, of course, in most simple amplifier interstage networks of the type already presented in this section the amplitude and phase responses are interrelated in a manner characteristic of the broad class of networks identified by the term *minimum phase*; it will not be possible here to go into the details of the definition of this terminology, nor into the details of the amplitude-phase relationship, but suffice it to say that feedback circuits, lattice and bridged-T structures, and distributed-parameter systems are the ones usually falling outside the minimum-phase class. A basic reference on the subject is Bode; see Reference 15. The relationship between amplitude and phase stems from a basic property involving the real and imaginary parts of a class of complex variables; see Reference 16. A brief discussion is also to be found in Terman; see Reference 17. Thus, it is somewhat futile to attempt to place the blame for transient distortion upon either the amplitude or phase distortion alone. Nonetheless, it is instructive at least to assess that distortion due to amplitude response, for it is possible—and indeed common practice on complicated transmission systems—to exploit a device known as a phase equalizer. By use of this device, which is an all-pass network of the nonminimum phase class, it is feasible to make the phase response more nearly linear, without influencing the amplitude response. The use of such equalizers is far beyond the scope of this treatment, although they are of great importance in long-distance television transmission systems. In fact, it might be said that their utility lies in systems that are limited to narrow frequency channels rather than in wideband systems where the gain-rise-time quotient of the amplifier tubes is the limiting factor. For references of interest on phase equalizers, see 18, 19, and 20. See also Bell System publications.

Special Amplifiers for High Speed

Although the details will be reserved for Section 24.4, it is appropriate to mention here that when fast amplifiers are required (amplifiers with small rise time T_R), faster than can be provided by the networks of Figures 24-4 through 24-19, there is the possibility of an altogether different kind of amplifier configuration. The circuits in the section above have been of the *product* variety, or *cascade*, the former term describing the fact that the overall gain function (of frequency) is the product of the individual stage gain functions, while cascade implies that one stage is connected after another. There is another class of amplifiers, which can be called *additive*, because of

the additive nature of their gain functions, and which permit of far greater speeds with conventional tubes than the cascade amplifier provides. This class includes the distributed amplifier and the *split-band* amplifier. Research is still being conducted in both these categories, but the present-day status will be discussed in Section 24-4.

24.3 Linear Video Amplifiers: Step Function Response (Sag, etc.) for Large Values of Time

The failure of practical amplifier circuits to transmit a step function perfectly for large values of time is in one sense a failure of the circuits to transmit direct current, or what is almost the same thing, very low frequencies. The long-time response to a step is generally better if the so-called low-frequency response of the amplifier is good, and indeed, most discussions of the subject in textbooks are phrased in terms of the low-frequency, steady-state behavior. The step response is often the desired criterion, however, and can be dealt with directly. For an example, in a television system, a picture in which there is a background (such as sky) with uniform intensity across the scene would require that the video amplifiers maintain constant (or almost) voltage for the duration of each horizontal scan period; thus, for an interval of about 60 μsec the step response must be constant to within some specified sag. The corresponding low-frequency behavior is purely incidental, and specification of the amplitude and phase response at low frequencies is at best indirect and not necessarily unique for a given sag.

It was shown in Section 24.1 that there are only certain portions of the complete circuit of Figure 24-2 that influence the long-time behavior; thus, the simplified circuit of Figure 24-6 adequately describes the response shown in Figure 24-7. This circuit contained only the coupling elements C_c and R_g, whereas in a practical pentode amplifier circuit it is necessary to consider also the cathode bias circuit (R_k and C_k in Figure 24-2) and the imperfectly bypassed screen-grid voltage supply. These will be taken up individually; the analysis follows closely that of Wallman (Reference 2, pages 84-92).

24.3.1 Coupling Circuit

This circuit has already been analysed for a single amplifier stage in Section 24.1, with results as shown in Figure 24-7. When a step function voltage is applied to the grid of one stage, the waveform delivered to the grid of the stage following is an exponential which decays to zero with time constant τ_g equal to $R_g C_c$.* This time constant is usually made very large compared to

*More exactly, $\tau_g = \left[\dfrac{r_p R_l}{r_p + R_l} + R_g \right] C$

the time interval within which the circuit must maintain a constant output voltage with a specified maximum sag. Thus, the circuit operation is confined in time to the initial portion of the exponential curve, which is closely a straight line having a slope, in percentage of the initial value, of $100/r_g$.

When several stages are connected in cascade, the time response is not a simple exponential, but will contain terms in t^2, t^3, etc., if the stages are identical or will be sums of exponentials if the stages are all different.

When the input signal is a one-volt step function, then in terms of a normalized time variable $t =$ (actual time)$/R_g C_r$, the results of Table 24-IV are obtained.

TABLE 24-IV

n	$e_m(t)$ Output Voltage	Initial slope, $t = 0$
1	e^{-t}	-1
2	$e^{-t}(1 - t)$	-2
3	$e^{-t}\left(1 - 2t + \dfrac{t^2}{2}\right)$	-3
4	$e^{-t}\left(1 - 3t + \dfrac{3t^2}{2} - \dfrac{t^3}{6}\right)$	-4

The situation regarding initial slope is of particular interest since this determines the sag. It can be seen from Table 24-IV that the slope (or sags) are additive. That is, the sag will increase directly with the number of stages. This is also true for the case of nonidentical stages.

24.3.2 Screen-Grid Circuit

The voltage supply for the screen grid in the usual pentode amplifier is obtained via a series resistor from the main plate-voltage supply, or occasionally from a voltage divider. In either case, the screen source has an internal resistance, and in order to limit the voltage variation at the screen there is customarily included a bypass capacitor from screen to cathode. Since the capacitor cannot be infinitely large, there is a voltage variation at the screen. This affects the space current in the tube, and hence the plate voltage is also affected.
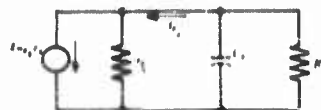
FIGURE 24-28
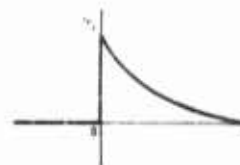
FIGURE 24-29

The first step in the analysis of this situation is computation of the screen current i_{v_y} (actually the incremental change in screen current, neglecting the steady d-c component). The equivalent circuit is as in Figure 24-28 and the waveform of i_{v_y} in Figure 24-29. Because of the proportionality between plate and screen current Figure 24-29 also depicts the waveform of plate current and load voltage. At the initial instant ($t = 0$) all the generator current $i_{v_y}(0)$ flows into the capacitor; hence:

$$i_{v_y}(0) = g_{m_{v_y}} e_v \tag{24-19}$$

Finally, as $t \to \infty$, the generator current will divide between r_{v_y} and R_s, such that:

$$i_{v_y}(\infty) = g_{m_{v_y}} e_v \frac{r_{v_y}}{r_{v_y} + R_s} \tag{24-20}$$

Time constant τ for the exponential charging of C_s:

$$\tau = \left(\frac{r_{v_y} R_s}{r_{v_y} + R_s} \right) C_s \tag{24-21}$$

Then the current slope at the initial instant is:

$$\frac{di_{v_y}}{dt} \bigg|_{t=0^+} = \frac{-g_{m_{v_y}} e_g}{r_s C_s} \tag{24-22}$$

The real interest is in current and voltage in the plate circuit, where the output of the amplifier stage is obtained. The following relationship can be employed:

$$\frac{i_p}{i_{v_y}} = \frac{g_m}{g_{m_{v_y}}} \tag{24-23}$$

Then the plate current slope becomes:

$$\frac{di_p}{dt} \bigg|_{t=0^+} = - \frac{g_m e_v}{r_{v_y} C_s} \tag{24-24}$$

Since the output voltage in the plate circuit is $i_p R_L$,

$$\text{Output voltage slope} = \left(\frac{-g_m e_g}{r_{g_p} C_s}\right) R_L. \tag{24-25}$$

Let $(g_m R_L) (e_g) = 100\%$. Then,

$$\text{Percent slope} = -100/r_{g_p} C_s \tag{24-26}$$

Cathode Bias Circuit

The third and last contributing cause of sag, or distortion in the long-time transient response, is the cathode bias circuit so frequently used in practical amplifiers to bias the grid sufficiently negative to prohibit the flow of grid current (Figure 24-2).

Qualitatively, what happens when a step function is applied to the grid is that the cathode bypass capacitor acts as a perfect short circuit at first, and

Figure 24-30

the tube amplifies with full gain. The capacitor slowly charges, however, and finally in the steady-state the bias resistor is essentially unbypassed. The effect of the cathode resistor is then degenerative, and the gain is reduced. The waveform is as in Figure 24-30.
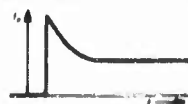
At the initial instant the bypass capacitor is completely effective, and

$$e_o(0) = (-g_m R_L)e_g \tag{24-27}$$

The final value is:

$$e_o(\infty) = -g_m R_L e_g/(1 + g_m R_k) \tag{24-28}$$

The time constant for the exponential decay is:

$$\tau = R_k C_k/(1 + g_m R_k) \tag{24-29}$$

From this the initial slope can be determined as:

$$\text{Slope} = \frac{e(\infty) - e(0)}{\tau} = (-g_m R_L e_g)\left(\frac{-g_m}{C_k}\right) \tag{24-30}$$

Letting $-g_m R_L e_g = 100\%$,

$$\% \text{ Slope} = -100\, g_m/C_k \tag{24-31}$$

Notice that the slope does not depend upon the size of the bias resistor R_k;

this is similar to the screen circuit, where the slope was independent of the dropping resistor R_n.

24.3.4 Summary

The three sources of sag; i.e., deficiency in the long-time transient response, give initial slopes as follows (for a single stage):

		(%/sec)
1.	Coupling Circuit (coupling capacitor C_p and grid resistor R_g)	$-100/R_g C_o$
2.	Screen Grid Circuit (dynamic screen resistance r_{g_b} and bypass capacitor C_s)	$-100/r_{g_b} C_s$
3.	Cathode Bias Circuit (grid-plate transconductance g_m and bypass capacitor C_k)	$-100 g_m C_k$

These relationships are valid only for small values of t/τ; i.e., for small enough time that the exponential response can be represented by a straight line.

The relationships were derived singly in order to ascertain which circuit elements governed the slope, and to obtain a quantitative measure of their contribution. In a practical amplifier the three effects will occur in combination in each stage, and in a cascade of stages the total effect will increase with the number of stages. In principle, to find the manner in which these effects should be combined one should examine a large number of cases or else provide some general formulation as Elmore has done for rise time in the short-time response. Lacking this, however, one can extrapolate from the results (shown in Table 24-IV) for a cascade of stages having only coupling-circuit deficiency. Here it was found that the initial slopes of $e(t)$ are directly additive as the number of stages is increased. Thus it seems not unreasonable to add directly the initial slopes of the time response.

$$\text{Total Slope (or sag)} \left. \right\} = \sum \text{ Individual Slopes (or sags)} \qquad (24\text{-}32)$$

From the evaluation of the separate causes of slope in the long-time response, it becomes evident which of them is the most severe. In particular it becomes apparent that the cathode-bias circuit is a worse offender than is the screen circuit; both depend only upon the tube characteristics and the

size of bypass capacitor, so a direct comparison can be made. As an example consider a 6AU6 tube, where typical values of g_m and $r_{g\,g}$ are 500 micromhos and 23 K respectively:

$$\text{Screen slope} = \frac{-100}{r_{g_g}C_s} = \frac{-100}{23,000C_s} = -0.004\%/\mu\text{sec }\mu\text{f}$$

$$\text{Cathode slope} = \frac{-100g_m}{C_k} = \frac{100 \times 5 \times 10^{-8}}{C_k} = 0.5\%/\mu\text{sec }\mu\text{f}$$

Thus, for the same slope in either case it would require a cathode bypass capacitor 100 times as large as the screen bypass capacitor. If the slope requirements are particularly stringent, it is sometimes expeditious to leave the cathode resistor unbypassed altogether. This reduces the gain, but removes the cathode slope entirely.

24.3.5 Cathode Peaking

In the special case where strict requirements on slope make it desirable to leave the cathode bias resistor unbypassed, it becomes necessary to re-evaluate the short-time transient response. In so doing it has been found beneficial to add a small capacitor across the bias resistor, instead of having none at all. This capacitance is so small, however, that its effect is negligible on the long-time response. In the short-time response, though, the rise time can be improved, with beneficial effects similar to those of shunt peaking, and hence the name *cathode peaking* is usually ascribed to this technique. The analysis proceeds from the circuit of Figure 24-31. The equivalent circuit is shown in Figure 24-32.
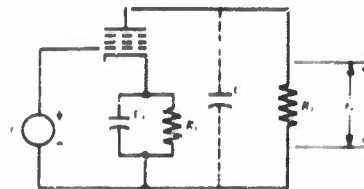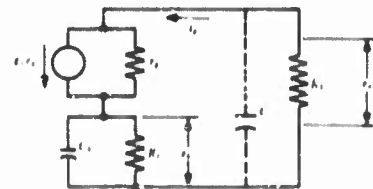
FIGURE 24-31   FIGURE 24-32

It will be convenient to define $\rho = R_k C_k / R_l C$ and $K = 1 + g_m R_k$. We are interested in trying different values of C_k for the circuit; the other elements are already determined. Thus, we can let ρ take values ranging from zero (cathode resistor unbypassed) to infinity (resistor perfectly bypassed). Three special values of ρ are of interest:

$$\rho = 0 \ (C_k = 0)$$

$$\frac{e_n(t)}{-g_m R_L} = \frac{1}{K}\left(1 - \epsilon^{-t/R_L C}\right) \tag{24-33}$$

$$\rho = \infty \ (C_k = \infty)$$

$$\frac{e_o(t)}{-g_m R_L} = \left(1 - \epsilon^{-t/R_L C}\right) \tag{24-34}$$

$$\rho = 1 \ (R_k C_k = R_L C)$$

$$\frac{e_n(t)}{-g_m R_L} =: \frac{1}{K}\left(1 - \epsilon^{-kt/R_L C}\right) \tag{24-35}$$
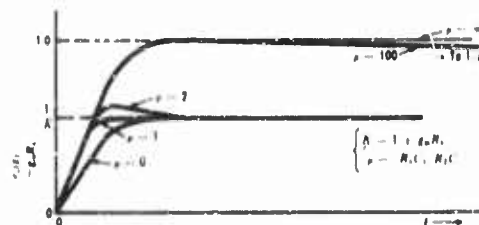
FIGURE 24-33

The time response for various values of ρ, and for a particular value of K (approximately 2), is shown in Figure 24-33. Notice that $\rho = 1$ gives a faster rise than does $\rho = 0$, and is the largest value of ρ that gives a monotonic rise (no overshoot); hence, this is the value usually chosen in the cathode-peaking technique. Notice also that $\rho = 100$ is a value in the usual range of cathode bypassing; there is a fast rise—essentially that of the ideal case $(\rho = \infty)$—followed by a long-time sag as previously analyzed for the cath-bias circuit.

It would appear from a comparison of Eq (24-34) and (24-35) that cathode peaking with $\rho = 1$ gives a faster amplifier than does the perfectly bypassed case. Both responses are simple exponentials with a 10 to 90 percent rise time already given in Eq (24-3):

$$\text{For } \rho = \infty \qquad \text{For } \rho = 1$$
$$T_R = 2.2 R_L C \qquad T_R = 2.2 R_L C/K$$

Thus, so long as K is greater than one, the rise time is indeed reduced by the factor K. Unfortunately, it is not usually sufficient to consider rise time alone. The peaking circuits in Section 24.2.2 were compared on their ability to improve the gain/rise-time ratio. On this basis the cathode-peaking circuit offers nothing, because the gain is reduced by the same factor that the rise time is reduced, namely K; hence, the ratio is the same as without peaking.

This perhaps conveys a needlessly unfavorable impression about cathode peaking. The reduction in gain can be restored by increasing R_L by a factor K. Then the gain is the same, the rise time is the same, but a much smaller bypass capacitor can be used, and there is no zag.

24.3.6 Slope Compensation

This is sometimes called low-frequency compensation, and rightly so, inasmuch as the slope or sag in the long-time response to a step is governed by those circuit elements which govern the low-frequency steady-state response. However, a steady-state analysis is perhaps at its worst when it comes to the question of how much compensation should be used for a good step response.

In transient terms we have already evaluated the long-time step response of the conventional circuit, and have found that sag results from several causes: coupling circuit, screen circuit, and cathode-bias circuit. Each of these is the source of a negative initial slope, and their effects tend to add directly. What we need is a circuit that will provide a *positive* initial slope, and one which will add to the negative slopes in the right amount to make the sum equal to zero.
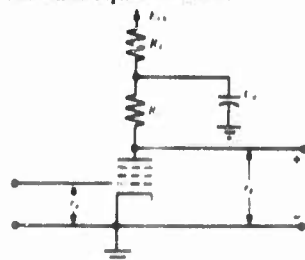
FIGURE 24-34

It turns out that the so-called decoupling circuit which is a normal part of a multistage amplifier—introduced to diminish coupling from one stage to the others via the common E_{bb} supply—is a suitable compensating circuit if proportioned properly. It is a long-time-constant circuit, the element values can be varied considerably without hampering its decoupling effectiveness, and it does provide a positive initial slope. The actual circuit is
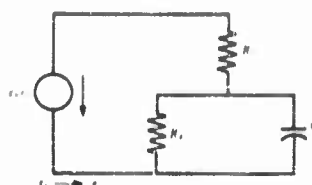
(A)

(B)

FIGURE 24-35

shown in Figure 24-34, and its equivalent for analysis purposes in Figure 24-35a, with the corresponding waveform of e_o for a step input e_g shown in Figure 24-35b.

In Figure 24-35 initial current $g_m e_g$ flows only through C_d (not R_d). Voltage rise across C_d is $g_m e_g / C_d$ and adds directly to $g_m e_g R_L$, which is initial value of e_p. Thus,

$$c_p = g_m e_g R_L + g_m e_g t / C_d \qquad (24\text{-}36)$$

$$\text{Slope} = + g_m e_g / C_c \qquad (24\text{-}37)$$

Let $g_m e_g R_L = 100\%$. Then $\qquad\qquad\qquad\qquad\qquad\qquad (24\text{-}38)$

$$\text{Slope} = 100\% / R_L C_d$$

The slope is evidently positive, and hence, by selecting a suitable value of C_o for the particular R_L being used it is possible to provide cancellation of the negative slope due to one or more of the causes discussed above. For instance, to cancel the slope due to the coupling circuit, choose C_d so that:

$$R_L C_d = R_g C_o \qquad (24\text{-}39)$$

By way of caution it should be pointed out that although R_d does not appear in the initial slope, it is necessary that $R_d C_d$ be large compared to the period of the step, both in order that the decoupling function be supplied and that the exponential function of Figure 24-35b be adequately approximated by the initial straight-line tangent.

24.4 Additive Amplification

Conventional amplifier systems, as discussed in the three sections preceding this one, could be called product amplification. This is to say, when several stages are connected in cascade, the over-all steady-state gain function (of frequency) is the continued product of the individual stage gain functions. In contrast, the amplifier structures to be described in this section have a gain function that is the sum of the "gains" provided by the separate elements.

Product, or cascade, amplification is older and more widely used. If the requirements on rise time (or bandwidth, in the steady-state) are not severe, a given amount of gain can be provided with fewer tubes in the cascade connection. But when the requirements are severe, it may be impossible to meet them with the product system, and yet the same tubes may be used in an additive structure to meet the requirements. The criterion that deter-

mines whether the cascade structure will work is the familiar quotient of gain over rise time, a factor that depends primarily on the tube (Table I), but which may be improved by more complicated circuits (e.g., Figure 24-21).

The difficulty becomes apparent when one is faced with providing a certain over-all gain for m stages (m unknown) and a certain over-all rise time T_{N_m}. There are various tube types available, each with a particular g_m/C, and one has at his disposal certain networks, each having an efficiency factor η expressing the relative speed with respect to the elementary resistance-coupled circuit. These are the building blocks. Unfortunately, there is a limit to what can be done. As given in Eq (24-11), the analysis by Elmore has shown that there is a minimum rise time that can be achieved with a given over-all gain. If the requirements call for a smaller rise time than Eq (24-11) permits, the job cannot be done with cascade amplification, except for the advent of better tubes or networks (the latter cannot be expected to provide an efficiency much better than four, and become inconveniently complicated at values of two).

Qualitatively, what happens is this. Suppose that a given rise time is stipulated, together with an over-all gain. One stage can be designed to provide this rise time, but it is found that the gain is insufficient. So, one goes to two stages, but in order to keep the over-all rise time the same, it is necessary to *reduce* the rise time of each stage. This is done by reducing the size of the load resistor, and hence reduces also the stage gain. If the over-all gain is still insufficient, the process is repeated. As the process is repeated, the stage gain becomes less and less, and would ultimately become less than unity. This would be an absurd state of affairs, because the over-all gain would then be less than the gain of one stage. The practical limit is actually reached when the stage gain has shrunk to 1.65, in accordance with Elmore's analysis.

Additive amplification is the way out. In fact, it even works when each tube contributes a gain of less than unity!

There are two principal types of additive structures. One of these is called the distributed amplifier; it was first employed in a British television installation in 1937, and has been in extensive commercial use in the U.S. since 1948. The other is the split-band amplifier, and is still under development.

Distributed Amplification

This form of amplifier structure was first proposed by Percival in 1935 in a British patent (Reference 21), although the system did not go into active use until after the first published analysis (Reference 22) in 1948.

The basic idea of the distributed amplifier is surprisingly simple, although there are, of course, many practical matters that contribute to the difference between actual performance and the first-order theory. The elementary form of the structure is given in Figure 24-36. The networks L_1C_1 comprise the so-called grid line, a cascade of filter sections in which the capacitors C_1 are the input capacitance of the tubes. Similarly, the networks L_2C_2 are the plate line,* C_2 being the output capacitance of the tubes. The two lines are designed to have the same phase velocity, and are terminated in their characteristic resistances R_{01} and R_{02}, respectively, so that no reflections take place. The lines are further assumed to be dissipationless, so that a wave can travel along either of them without attenuation.

Within the limits of these idealized conditions, the following relationships hold:

$$\text{Characteristic impedance} \quad \begin{cases} R_{01} = \sqrt{L_1/C_1} \\ R_{02} = \sqrt{L_2/C_2} \end{cases} \qquad (24\text{-}40)$$

$$\text{Phase velocity (sections/sec)} \quad v_p = 1/\sqrt{L_1C_1} = 1/\sqrt{L_2C_2} \qquad (24\text{-}41)$$

$$\text{Plate current of each tube} = g_m|E_1| \qquad (24\text{-}42)$$

Because of the equal velocities in grid and plate lines the plate current contributions of successive tubes will add directly, i.e., they are all in phase at the load.

$$\text{Load Current} = n\, g_m E_1/2 \qquad (24\text{-}43)$$

The factor 2 in Eq (24-43) comes in because half the current contributed by each tube flows to the left in the plate line, and is lost in the terminating resistor R_{02}. In spite of the current thus lost, the resistor R_{02} is usually necessary to prevent reflections due to the wave travelling to the left, which at certain frequencies could cancel the wave travelling the right at the load.

$$\text{Output voltage,} \quad E_o = n\, g_m E_1 R_{02}/2 \qquad (24\text{-}44)$$

$$\text{Amplification,} \quad A \overset{\triangle}{=} E_o/E_1 = n g_m R_{02}/2 \qquad (24\text{-}45)$$

Equation (24-45) displays the basic property of the distributed amplifier.

*The conventional filter formulas are open to question where the plate line is concerned, since the structure is driven by a number of current generators along the structure instead of by a single generator at one end. This situation is analyzed in References 23 and 24.

namely that the amplification increases linearly with the number of stages, i.e. each tube contributes a gain of $g_m R_{02}/2$, and the total gain is the *sum* of the individual contributions. Indeed, each tube may contribute a gain of less than unity, and yet the total gain can be made as large as desired by adding a sufficient number of tubes. This situation is not possible with cascade amplification.

It is feasible—and frequently advantageous—to cascade stages of the type of Figure 24-36. This is done by connecting the grid line of the second stage
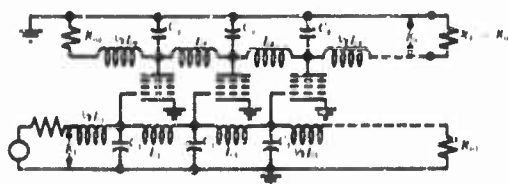
FIGURE 24-36

as the load on the first stage (with a blocking capacitor for d-c isolation, of course). Should R_{01} and R_{02} not be the same, a transformer is in principle required to join the two stages in order to prevent reflection. A transformer with sufficient bandwidth (up to 200 mc in typical distributed amplifiers) is not available, and so it is more usual to make R_{01} and R_{02} the same by adding to the smaller of C_1 or C_2 in order to equalize them. The effect is the same, though; the gain is reduced. With the transformer, (24-45) becomes:

$$A = n g_m \sqrt{R_{01} R_{02}/2} \qquad (24\text{-}46)$$

If, instead of a transformer, capacitance is added to the smaller C:

$$A = n g_m (R_{01}/2) \quad \text{if} \quad R_{01} < R_{02} \qquad (24\text{-}46a)$$

The question to be asked about cascading is, when does one stop adding tubes along the line in one stage, and add further tubes in a second stage? This readily is answered for the idealized conditions we have thus far considered, i.e., that the grid and plate networks behave perfectly as lines. Suppose that a total gain A, is required, how few tubes, N, are necessary in m stages, with n tubes per stage? The contribution of each tube to the stage gain A in Eq (24-46) is A_t, where

$$A_t = (g_m/2) \sqrt{R_{01} R_{02}}$$

therefore,

$$A = n A_t$$

and

$$A_t = (nA_i)^m \qquad (24\text{-}47)$$

also

$$N = nm \qquad (24\text{-}48)$$

From the equations (24-47) and (24-48) it is possible by differentiation of N with respect to m to find a minimum value of N for a given A_t. This minimum occurs when

$$A = nA_i = \epsilon$$
$$= 2.72 = 8.7 \text{ db} \qquad (24\text{-}49)$$

This relationship is given by Ginzton, *et al.*, and by Copson in Reference 25. The relation assumes that there is no bandwidth shrinkage as stages are cascaded, which is correct for narrow, low pass cascades. In very wide-band amplifiers the frequency response diminishes with attenuation due to loading of the grid line by the input conductance of the tubes. Allowing for bandwidth shrinkage according to an arbitrary function

$$B_m/B_i = 1/\sqrt{m}$$

gives an optimum stage gain of $\epsilon^{2/3}$ instead of ϵ as in Eq (24-49).

Thus, the most efficient use of tubes results from cascading stages of no more nor less than 8.7 db per stage. Of course, we have learned that cascading stages impairs rise time and reduces bandwidth, and this consideration has been neglected here.

The analysis can proceed no farther without a study of the grid and plate networks. The basic premise in the elementary theory of the distributed amplifier is that these networks simulate smooth lines, terminated in the appropriate R_0 so that no reflections occur. Apart from the termination problem, the important requirement is that the velocity be constant with frequency. This is the same as saying that the phase shift must be proportional to frequency. Failure to realize this results in *phase distortion*, which, as was considered in Section 24.2, shows up also in distortion of transient signals. Naturally, it is true that the *amplitude distortion* (as a function of frequency) also enters into the transient distortion. It appears, however, that the elementary ladder networks such as in Figure 24-36 have a disproportionate amount of phase distortion, and hence an all-out effort to improve this results in a better balance between amplitude and phase distortion, and a better transient response.

The problem of designing the best possible network for use in the distributed amplifier has its counterpart in the design of the best possible delay line from lumped, dissipationless elements. These delay lines are used in oscilloscopes, radar, etc., and provide small time delays with as little distortion as possible, i.e., distortion of rectangular pulses or other transient waveforms. Interestingly enough, the network most widely used in delay lines is the same as that in the typical distributed amplifier. References on the delay line include 26, 27, 28 and 29. Actually, the distributed amplifier presents a somewhat more difficult network problem, because the input conductance of the tubes connected to the grid line introduces dissipative elements in shunt with the capacitances; this will be discussed later.

Consider now the basic circuit of Figure 24-36. The grid and plate networks have the appearance of transmission lines in which the infinitesimal elements of series inductance and shunt capacitance in the continuous line have become finite. It is, indeed, an artificial line, and the properties of such lines (in the steady-state) have been known since the early days of the telephone. Moreover, the whole art of the wave-filter was first devised by G. A. Campbell from the concept of the artificial line, and we can profitably use the results of much wave-filter analysis in the problem here.

The artificial line fails us because it has one characteristic not possessed by the smooth line, namely a *cutoff frequency* f_o, where

$$f_o = 1/\pi\sqrt{LC} \qquad\qquad (24\text{-}30)$$

The significance of the cutoff frequency for the network structure in Figure 24-36 (which is shown again in both its T and pi equivalents, in Figure 24-37
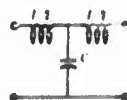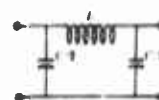
FIGURE 24-37   FIGURE 24-38

and 24-38, respectively) is that a chain or ladder of sections of this type will transmit frequencies below, but not above, this frequency. Moreover, there is an abrupt discontinuity in the phase velocity and the characteristic impedance* at the cutoff frequency, and indeed these parameters begin to vary significantly with frequency long before the cutoff is approached. Equations

*In filter analysis this is called the *image impedance*, which in the case of the input impedance to an infinite ladder of sections is also the characteristic impedance. The latter is defined in terms of reflections, whereas the former is defined as the input impedance to a section when terminated at the outlay by an impedance equal to the impedance looking back into the output terminals, i.e. the "image", using the mirror analogy.

(24-40) and (24-41) for the characteristic impedance and velocity, are those that would be appropriate for a smooth line, if L and C represented the values per unit length; they also hold for the lumped network at very low frequencies (far from cutoff). The general expressions for the impedance and velocity at any frequency in the region below cutoff are:

$$\text{Characteristic (image) impedance } R_0 = \frac{\sqrt{L/C}}{\sqrt{1 - (f/f_o)^2}} \qquad (25\text{-}51)$$

$$\begin{matrix}\text{Phase velocity} \\ \text{(sections per second)}\end{matrix} \quad v_\theta = \frac{1}{\sqrt{LC}} \sqrt{1 - (f/f_o)^2} \qquad (24\text{-}52)$$

The delay τ per section is defined as $1/v_p$. A plot of τ and R_0 is given in Figure 24-39.

$$\tau = \sqrt{LC} / \sqrt{1 - (f/f_o)^2} \qquad (24\text{-}53)$$

It can be seen from the figure that there are two problems which were not considered in the ideal line case. One is the termination problem, and the other the time delay, or velocity problem. It is important that the grid and plate networks be terminated so that there are no reflections; otherwise, as the frequency is varied, the reflected wave could alternately add to and subtract from the forward wave, thus producing variations in the amplifier gain as a function of frequency. From Figure 24-39 it is apparent that the terminating impedence must vary with frequency in the manner indicated there, instead of being a ample resistor as depicted in Figure 24-36.

The other problem, time delay, is serious from the standpoint of the transient response. The curve of Figure 24-39 displays phase distortion, i.e. the time delay is not constant as would be the case with phase proportional to frequency. The discussion in Section 24.2 brought out the undesirable transient performance resulting from phase distortion.
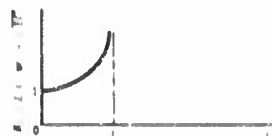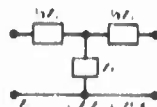
FIGURE 24-39   FIGURE 24-40   FIGURE 24-41   FIGURE 24-42
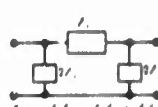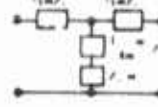
Both of these problems can be solved sufficiently well for practical purposes, although the solutions are by no means the ultimate. Consider first the termination problem. The proper terminating impedance, with characteristics as in Figure 24-39, can be provided by what the filter people call an

m-derived half-section. A few words might be in order concerning the background of terminology, as well as the device itself.

The network sections of Figures 24-37 and 24-38 are low-pass forms of general filter sections shown in Figures 24-40 and 24-41. Proper choice of the elements making up Z_1 and Z_2 permits not only low-pass structures, but also high-pass, bandpass and band-elimination filters. One basic type of all these is the so-called constant-*k* structure, in which, independent of frequency, the following relationship holds:

$$k^2 = Z_1 Z_2$$

The *m*-derived section evolves from the constant-*k* one in the following manner. If a network section is assembled as in Figure 24-42, in which Z_1 and Z_2 are the same as in the constant-*k* (sometimes called "prototype") section of Figure 24-40, such a section is said to be *m*-derived. The coefficient *m* can be any real constant, not necessarily an integer. The *m*-derived counterpart of Figure 24-37 is shown in Figure 24-43.

The *m*-derived section of Figure 24-43 has the same cutoff frequency and the same variation of image impedance as does the prototype in Figure 24-37. Hence, such a section could be inserted in a ladder of prototype sections without producing reflections at any frequency. There are positive advantages to
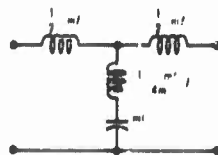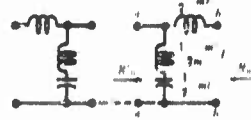
FIGURE 24-43   FIGURE 24-44   FIGURE 24-45
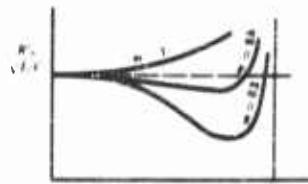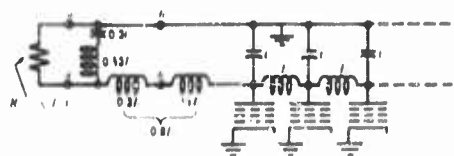
FIGURE 24-46   FIGURE 24-47

the *m*-derived section, beyond this permissive attribute of being able to use the section in combination with the termination problem. Suppose that we split the *m*-derived section of Figure 24-43 into two half sections as in Figure 24-44, and then examine the image impedance looking into the terminals

a-a; this is called the mid-shunt impedance. For various values of m it varies with frequency as indicated in Figure 24-45. This is quite remarkable, because for *all* values of m the image impedance of the main section (terminals b-b) remains as in Figure 24-39. The value $m = 0.6$ is particularly useful. Notice that $R'_0/\sqrt{L/C}$ is very close to 1.0 to a frequency nearly equal to f_o. Our terminating technique is now in hand. We connect a resistance $\sqrt{L/C}$ to the terminals a-a; this matches the image impedance R'_0 (within the limits of the approximation of the $m = 0.6$ curve), and hence looking to the left at b-b the impedance R_0 is the same as though there were an infinite number of either m-derived or constant-k sections extending to the left of a-a. Then we connect the terminals b-b to the ladder extending to the right, which ladder is then terminated. In Figure 24-4 this is done for the plate network of the distributed amplifier.

Our first problem is thus solved, with the aid of the terminating half-section. Four of these terminations are usually required, one at each end of both grid and plate networks.

The next problem, that of the frequency-variable time delay, is also solved with the employment of an m-derived structure. If one explores the time delay per section for various values of m, curves such as those of Figure 24-47 are obtained. Notice that $m = 1.27$ has a particularly favorable characteristic, in that the delay time remains approximately constant to a high frequency; this is the value usually chosen for delay lines and distributed amplifiers. Remember that choosing any particular value of m, such as 1.27, has no effect on the image (characteristic) impedance, so all that has been said about terminations, etc., is unaffected by whether we use constant-k or m-derived sections associated with each tube in our amplifier.

There is one slightly embarrassing feature in choosing a value of m greater than unity: a negative inductance is required in the shunt branch (see Figure 24-43). Such a requirement can be met, however, by providing mutual inductance of the proper amount between the two series elements. Thus, in Figure 24-43 we interpret the three inductances to represent the T equivalent of a transformer, and then replace the equivalent by an actual transformer. It is possible for the actual transformer to be physically realizable, even though the T equivalent is not. Thus, the circuit of Figure 24-43 becomes that of Figure 24-48. In practice the transformer is constructed by tapping onto a single layer coil of suitable proportions (see Kallman papers and Reference 30), as indicated in Figure 24-49.

Thus, each section of the distributed amplifier is arranged as in Figure 24-49, where the capacitance is either the input or output capacitance of the tube. The complete form of the plate network for instance, would then be as in Figure 24-50.
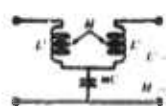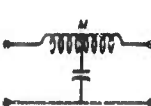
FIGURE 24-48   FIGURE 24-49   FIGURE 24-50
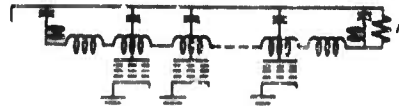
This, then is the basic design philosophy of the distributed amplifier, as currently being used. The steady-state response is quite acceptable to very high frequencies, in spite of the fact that the approximation involved in the terminating half-sections gets worse as the cutoff frequency is approached. From the standpoint of transient response, the approximation does not seem to work out as well, and in practice there is a certain amount of cut-and-try manipulation of the termination on an experimental basis, while the operator observes the shape of the transient response.

Other forms of networks have been employed, such as the bridged T, but the arrangement of Figure 24-50 is the most widely used. It is possible, for instance, to use a continuous solenoid for the plate or grid line, with taps along its length for the tube connections (Reference 31).

In a practical case, where the amplifier is to operate at frequencies of one or two hundred megacycles, the analysis should be extended to include the effects of the input admittance of the pentode tubes. This input admittance has both a capacitive and a conductive component. The latter is the more serious in the distributed amplifier, because it produces attenuation of the signal travelling down the grid network. The conductance of the tubes increases with the square of frequency, and so as the frequency is increased the attenuation ultimately reaches a level such that the attenuation per section is greater than the gain provided by the tube; beyond this frequency the gain diminishes—and adding more tubes only makes matters worse.

The attenuation due to input conductance has been analyzed in the literature (Reference 32). One important conclusion is that, since the magnitude of the input conductance is proportional to the cathode lead inductance and cathode-grid transit time of the tube, the choice of tube must involve these factors as well as the usual ratio of g_m to capacitance.

A question which might well be asked is, when does one change from using cascade amplification in favor of distributed amplification? A partial answer is given by the graph of Figure 24-51, taken from the paper by Ginzton, et al. Here is plotted the number of tubes required to give a gain of ϵ, which is the optimum stage gain in the distribution amplifier, i.e., if more gain is needed, stages should be cascaded. This value of gain, while low, is close to the value $\sqrt{\epsilon}$, which came from Elmore's analysis as the optimum stage gain for rise time. The independent variable in the graph is bandwidth ratio, or in other words a ratio of high-frequency cutoffs. It is the ratio of the
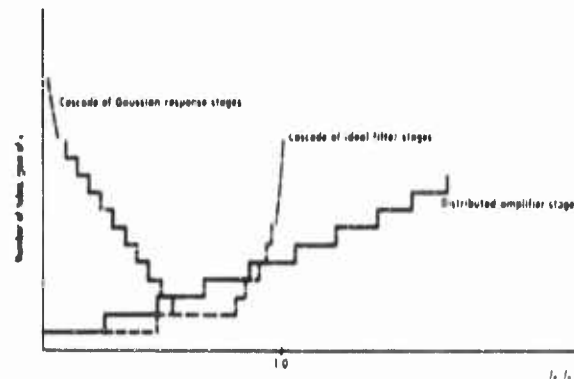
FIGURE 24-51

desired high-frequency limit f_a of the amplifier to the high-frequency limit of the tube, f_o. Two curves are shown for the cascade amplifier. With ideal filter stages it is assumed that there is no narrowing of the bandwidth as the number of stages is increased, while with a gaussian-type response, the narrowing is assumed to be like that of RC stages but with a gain-bandwidth factor of 4.

The conclusion from Figure 24-51 is that for small bandwidths, i.e., small f_a/f_o (or what is equivalent, large rise time), the cascade amplifier is better, whereas for very large bandwidths, the distributed amplifier is more efficient. Indeed, if the bandwidth requirement is greater than that corresponding to an f_a/f_o greater than unity, *only* the distributed amplifier can do the job.

Split-Band Amplifier

The split-band amplifie   sometimes called the parallel chain amplifier or

FIGURE 24-52

FIGURE 24-53

divided-band amplifier, has been proposed at various times, and is the subject of a limited amount of current research (References 33 and 34). The basic structure consists of two (or more) amplifiers in parallel, each providing gain over a portion of the entire passband needed, as depicted in Figures 24-52 and 24-53. The individual amplifiers may be either cascade or distributed. Although the ultimate performance of this split-band structure holds great promise, there is difficulty in designing the branching networks N_1 and N_2, and the characteristics of each amplifier, so that the entire assembly has the desired frequency response, particularly in the critical crossover region.
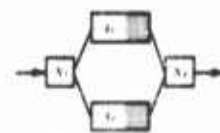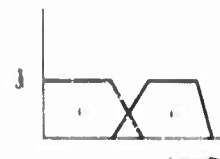
24.5 Introduction to the Filter Amplifier: I-F Amplifiers

The discussion of video amplifiers centered around the transient response. The function of the video amplifier was to provide gain with as little distortion of the waveform as possible. In many systems, however, such as ECM receivers, there is a different kind of requirement. Not only is an amplifying system expected to produce gain over a bandwidth sufficient for the information contained in the signal, but it must also *reject* signals outside the appointed band. This is primarily a steady-state matter; we are talking in terms of the frequency response of the system. It may of course be true that the desired signal which is to be amplified consists primarily of transient waveforms, and that the transient response must therefore not be entirely ignored. Nonetheless, we shall seek to allow for this through use of the interrelationships described in the preceding sections, between steady-state and transient response.

Our principal concern will be with intermediate-frequency (i-f) *bandpass* amplifiers, rather than with lowpass or highpass. In one sense we considered the lowpass amplifier in the preceding sections, but there were no requirements that the amplifier reject frequencies above any prescribed frequency limit. Bandpass amplifiers are fundamental to all types of systems using the radio-frequency spectrum on a frequency-separation basis (as opposed to time separation, e.g. multichannel PTM), including ECM, radio, radar, and carrier on wire lines.

There are several aspects of the design requirements, and their relative importance will vary from one application to another. These are the gain magnitude (band center), the bandwidth, and the center frequency. It shall be the objective here to present techniques which make possible the achievement of really difficult combinations of requirements; viz, high-gain, large bandwidth, and high center frequency, all in one amplifier, together with good rejection of signals outside the passband.

High-gain, bandpass amplifiers are usually to be found in the receiver portions of systems, where the signals are of small amplitude. Hence we can deal with the small-signal equivalent circuit of the tubes, which in most cases will be pentodes (Figure 24-54).
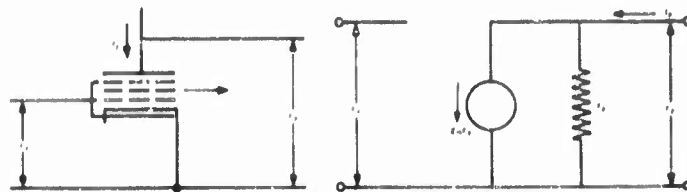
FIGURE 24-54
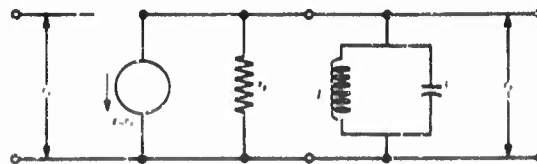
FIGURE 24-55

From the figure it is perceived that the tube provides a *linear* (constant g_m), unilateral network, associated with which will be a group of circuit elements comprising an *interstage coupling network* (referred to later simply as an interstage). The two in combination comprise an amplifier *stage*. An example is shown in Figure 24-55, in which the interstage is simply a parallel resonant circuit.

It will turn out that in order to achieve large bandwidths there will be added no additional shunt capacitance beyond that provided by the input and output capacitances of the tube, C_i and C_o, respectively. Moreover, at high frequencies the effects of transit time and the inductance of the cathode lead in the tube contribute to an appreciable *input conductance* g_i. In these circumstances the equivalent circuit of the tube appears as in Figure 24-56. It should be pointed out, however, that the input conductance really varies with frequency (increasing as f^2), but is usually considered constant over the pass band of the amplifier in practical cases.
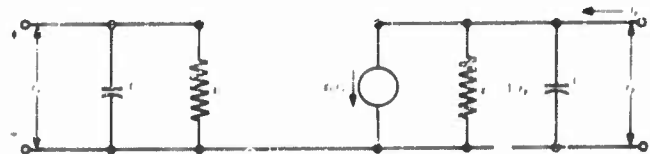
FIGURE 24-56

Even the circuit of Figure 24-56 is still a linear, unilateral network. The unilateral property would be destroyed if the grid-plate capacitance of the tube were significant, but in most modern, high-gain pentodes this capacitance can be ignored.

Our basic problem consists in assembling useful combinations of tubes and interstages, usually in the cascade (or "tandem," or "chain") arrangement, in such a way and with such components as to provide a desired gain and amplitude response. (Note: *amplitude response* will be the abbreviated way of saying "amplitude response versus frequency," or "gain magnitude versus frequency.")

The problem is similar to that of designing a filter consisting of passive network elements only. Indeed the properties of the over-all transfer func-

tion of an amplifier have many points in common with that of the passive filter. It might be said in this regard that we are dealing with applied network theory. Our task is more difficult in two respects: (1) the amplifier must provide gain, and (2) we are constrained to interstage networks that will function properly in combination with the irreducible equivalent net..ork of the tubes (e.g., Figure 24-56). As a consolation, however, it will turn out that the unilateral property of the tubes serves to "isolate" the interstage networks, and certain advantages are gained. both in the design and alignment of the amplifier, as contrasted with the completely passive structure.

Since the interstage networks can be only those which "fit in" with the tube networks, it might well be expected that only certain classes of interstages have been found to have practical value. This is indeed the case, and the *synthesis* of the filter amplifier is largely a matter of *selection*, both as to tube and network, based upon suitable *figures of merit* for comparing their value. Therefore we shall commence the study by an examination of several types of interstage network, beginning with the simplest.

We shall not proceed far in the direction of *network* complexity, although the frequency response of a multistage amplifier may be identical in its characteristics to that of a complicated filter. Where many tubes are needed to supply *gain*, it is good strategy to distribute nonidentical, simple filter-type networks between the tubes, all designed in such a way as to yield the desired over-all response. This comprises the so-called filter amplifier, a term introduced by Butterworth (Reference ₃5).

When only a few tubes may be needed to provide the required gain, and yet the desired frequency response be that of a filter with many sections, it is possible to use networks of greater complexity than those commonly used with filter amplifiers. Such networks differ from the usual passive filters in that they must provide for capacitance at input and output terminals and also for finite () of the inductors (Reference 36). Adjustment of such networks calls for special techniques (Reference 37).

24.6 Two-Terminal Interstage Network: The "Single-Tuned" Interstage

24.6.1 Properties of a Single Stage

As a starting point, suppose that our gain requirements are so modest as to permit the use of one stage alone. Or, even if they are not. the single stage is the fundamental "building block" in the multistage amplifier.

Next, let us see what can be done with the simplest possible interstage network, and use the performance of such a network as a calibration for comparing other more complicated networks.

As a practical ma   , the only form of two-terminal interstage in extensive use today is the single-tuned network.

Starting with the basic equivalent network for the tube (Figure 24-26) the single-tuned interstage is constructed by adding a shunt inductance L, tuned to resonance at the desired center frequency with the total interstage capaci-
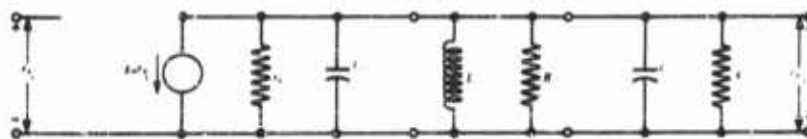
FIGURE 24-57

tance $C_i + C_o$, and a resistance R. This is illustrated in Figure 24-57.

With most pentodes it is possible to ignore r_p and g_i in comparison with R, in which case the circuit reduces to that of Figure 24-58. In any event, R can represent the parallel combination which includes r_p and g_i. The capacitor C likewise represents the parallel combination of C_o, C_i and stray wiring capacitance.

FIGURE 24-58   FIGURE 24-59

It must be pointed out here that in adding parallel R we have elected to explore the wideband case. If we wanted a narrow-band amplifier, we would add no shunting resistance, and would be careful that the inductor L had the least possible losses (highest possible Q). The inductor losses, usually considered as a series resistance R_s, are generally the dominant factor in determining the bandwidth, and hence the proper equivalent circuit is that of Figure 24-59. This case is thoroughly explored in the radio engineering books, (for example, see Reference 1) and we leave it now in favor of the more difficult task of maintaining a wideband simultaneously with high gain.

We proceed with the analysis, after first defining some terms. The *gain function* $A(j\omega)$ is the complex number which contains the amplitude and phase of the ratio $E_{y_g}(j\omega)/E_{p_1}(j\omega)$. The *gain* (without the word "function") is taken to be the midband gain, i.e., $|A(j\omega_0)|$, where ω_0 is the radian frequency at the center of the band. Finally, the *bandwidth B* is arbitrarily defined as the interval between the two frequencies (one higher and one lower than the center or midband frequency) at which the magnitude of the gain function $|A(j\omega)|$ is down to 70.7% (-3 db) of the midband gain.

For Figure 24-58:

$$A(j\omega) = \frac{E_{g_2}}{E_{g_1}} = \frac{-g_m E_{g_1} Z(j\omega)}{E_{g_1}} \tag{24-54}$$

where $Z(j\omega) = 1/(1/R + j\omega C + 1/j\omega L)$ for

$$Q = R/\omega_0 L \qquad \text{and} \qquad \omega_0 = 1/\sqrt{LC} \tag{24-55}$$

$$A(j\omega) = -g_m R \frac{1}{1 + jQ\left(\dfrac{\omega}{\omega_0} - \dfrac{\omega_0}{\omega}\right)}$$

$$= -g_m R \frac{1}{1 + jQ\left(\dfrac{f}{f_0} - \dfrac{f_0}{f}\right)} \tag{24-56}$$

where $f = \omega/2\pi$
  $f_0 = \omega_0/2\pi$

Maximum gain at midband ($f = f_0$):

$$|A(j\omega_0)| = g_m R \tag{24-57}$$

Bandwidth limits determined by f to give:

$$A(j\omega) = -g_m R \frac{1}{1 + j1}$$

$$|A(j\omega)| = |A(j\omega_0)| \frac{1}{\sqrt{2}}$$

or

$$\frac{f_1}{f_0} - \frac{f_0}{f_1} = \frac{f_0}{f_2} - \frac{f_2}{f_0} = \frac{1}{Q}$$

$$\text{Bandwidth } B = f_1 - f_2 = f_0/Q \tag{24-58}$$

$$= 1/2\pi RC \tag{24-58a}$$

The two frequencies f_1 and f_2, at which the *magnitude of the gain function* (sometimes called the *amplitude response*, or simply response) is down 3 db from that at band center f_0, are related by:

$$f_0/f_1 = f_2/f_0 \tag{24-59}$$

Thus, f_0 is the geometric mean of f_1 and f_2, and the passband of the amplifier is said to possess "geometric symmetry" about f_0.

Notice that in Eq (24-58a) the bandwidth is independent of the center frequency f_0. (In (24-53) f_0 appears, but it is also implicit in Q and cancels out in the quotient.)

Notice also that in Eq (24-57) the gain is independent of the center frequency. Most important of all, however, observe that in the product of (24-58a) and (24-57), the gain-bandwidth product is not only independent of frequency but is also independent of the resistance R:

$$\text{Gain} \times \text{Bandwidth} = g_m/2\pi C \qquad (24\text{-}60)$$

This shows that gain can be exchanged for bandwidth (by varying R), but the product remains constant, and is independent of center frequency.*

The product of gain times bandwidth for the single-tuned circuit as given by (24-60) is determined primarily by the tube. Except for stray wiring capacitance, the quantity C is largely due to the tube, and so also, of course, is g_m. This tells us that in selecting a tube for large gain and bandwidth, the g_m/C ratio should be as large as possible. For a chart showing the g_m, C, and g_m/C for currently available commercial tube types, see Reference 38. Moreover, there is no benefit in a large g_m—as might be achieved by connecting several tubes in parallel—if the capacitance C is increased in the same proportion.

The expression in Eq (24-60) is the gain-bandwidth product, a term which we shall distinguish carefully from the gain-bandwidth *factor*. The latter is a figure of merit that is useful in comparing various interstage *networks* against the single-tuned stage as a "yardstick". This is what we did in the transient analysis of fast amplifiers; the relative speed of various circuits such as shunt peaking was compared to the resistance-coupled stage. Thus, in the case at hand, the gain-bandwidth *factor* of a circuit is obtained by taking the gain-bandwidth *product* and dividing it by the gain-bandwidth product of the single-tuned stage, namely $g_m/2\pi C$. We shall find that for two-terminal interstages the maximum gain-bandwidth factor is 2.0, while with four-terminal networks it can theoretically be as great as $\pi^2/2$ or 4.93.

The gain function $A(j\omega)$ given in Eq (24-56) can be simplified if the circuit bandwidth is, say, 10 percent or less.

*It is worth noting that there is a practical limit on how high the center frequency may be, determined by the input conductance g_i, which increases as f^2. If the load resistor is removed, and $R \to 1/g_i$, then the gain given by Eq (24-57) becomes g_m/g_i. This quantity will diminish with increasing frequency until at some limiting frequency the stage gain will become less than unity. For example, a 6AK5 may have $g_m = 5 \times 10^{-3}$ and $g_i = 0.14 \times 10^{-7} f^2$; the gain will thus drop to unity when f reaches 190 mc.

$$\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \cong \frac{\omega^2 - \omega_0^2}{\omega_0\omega} \cong \frac{(\omega + \omega_0)(\omega - \omega_0)}{\omega_0\omega}$$

$$\cong \frac{2(\omega - \omega_0)}{\omega_0}$$

Narrow Band:

$$A(j\omega) = -g_m R \frac{1}{1 + j\frac{2Q}{\omega_0}(\omega - \omega_0)} \qquad (24\text{-}61)$$

$$= -g_m R \frac{1}{1 + j\frac{2Q}{f_0}(f - f_0)} \qquad (24\text{-}61a)$$

Notice that in this narrow-band case, the function has "arithmetic symmetry" about ω_0; i.e., the two frequencies ω_1 and ω_2, at which the response is down to 70.7 percent are equally displaced in frequency increment above and below ω_0, or:

$$\left. \begin{array}{l} \omega_1 - \omega_0 = \omega_0 - \omega_2 \\ f_1 - f_0 = f_0 - f_2 \end{array} \right\} \qquad (24\text{-}62)$$

24.6.2 Multistage Single-Tuned Amplifier

It would be unusual if the gain of one stage were adequate; most applications require several stages in cascade. The intermediate-frequency amplifier in a radar set might have as many as ten stages, while a transcontinental relay link would have hundreds of stages.

In cascade amplification the gain functions combine in a continued product. The midband gain will be the nth power of the stage gain; namely,

$$|A(j\omega)| = (g_m R)^n \qquad (24\text{-}63)$$

Simultaneously, the bandwidth will decrease as the number of stages is increased. Thus the -3-db points for one stage become the -6-db points for two stages in cascade, with a resultant smaller separation of the -3-db points for the pair. To solve analytically for the manner in which the bandwidth "shrinks" it will be convenient to rewrite Eq (24-56) and (24-61) in a normalized frequency variable x as follows:

$$\text{Selectivity Function} = F(jx) = \left(\frac{1}{1 + jx}\right)^n \qquad (24\text{-}64)$$

In terms of this normalized variable the bandwidth of a single stage is 2.0, and against this we can compare the bandwidths for larger numbers of stages.

$$|F(jx)| = \frac{1}{\sqrt{1 + x^2}} \qquad x \overset{\Delta}{=} Q\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad (24\text{-}65)$$

For $\frac{1}{2}$ bandwidth x_n,

$$\frac{1}{\sqrt{1 + x_n^2}} = \frac{1}{\sqrt{2}} \qquad (24\text{-}66)$$

$$x_n = \sqrt{2^{1/n} - 1} \qquad (24\text{-}67)$$

For $n \gg 1$,

$$x_n \approx \frac{0.833}{\sqrt{n}} \qquad (24\text{-}67a)$$

Now, since the bandwidth for one stage is 2.0, the half bandwidth is 1.0, and hence Eq (24-67) is also the ratio of the bandwidth of n stages compared to the bandwidth of a single stage:

$$\frac{\text{Bandwidth of } n \text{ stages}}{\text{Bandwidth of one stage}} = \sqrt{2^{1/n} - 1} \qquad (24\text{-}68)$$

$$\approx 0.833/\sqrt{n} \qquad (24\text{-}68a)$$

for $n \gg 1$

Thus, in a cascade of single-tuned stages the gain increases with the number of stages according to one law, Eq (24-63); while the bandwidth varies according to another; namely, Eq (24-68). Thus the contribution of one stage to the over-all gain-bandwidth product will be different if the stage is used alone or in a cascade with other stages. To allow for this in our figure of merit, the *gain-bandwidth factor* (GBF) by means of which we shall compare the effectiveness of various circuits we do two things: (1) take the nth root of the over-all gain, giving a *mean stage gain*, (2) divide the mean stage gain by $g_m/2\pi C$. The gain-bandwidth factor is then the product of this normalized mean stage gain times the over-all bandwidth. It is a function of the circuits used, rather than of the tubes.

$$\text{GBF} = \frac{\text{Gain-Bandwidth}}{\text{Factor}} = \frac{(\text{Over-all Gain})^{1/n}}{g_m/2\pi C} (\text{Over-all Band-width}) \quad (24\text{-}69)$$

A phenomenon of both mathematical and practical interest is that, for a given bandwidth, there is a definite maximum of gain which cannot be exceeded. There is likewise a maximum bandwidth attainable for a specified gain. The latter occurs when each stage contributes a gain A_i of 1.65 to the over-all gain A_n. The number of stages needed and the resulting maximum bandwidth B_{max} are given below:

$$A_i = \sqrt[n]{e} = 1.65 = 4.34 \text{ db} \qquad (24\text{-}70)$$

$$n = 2 \ln A_n \qquad (24\text{-}71)$$

$$B_{max} = \frac{g_m}{2\pi C} \frac{0.833}{\sqrt[n]{2e \ln A_n}} \qquad (24\text{-}72)$$

24.6.5 Selectivity Ratio

It was stated that the filter amplifier has the dual objective of amplifying signals within the desired passband, and of rejecting signals lying outside this band. The ability to reject unwanted signals is sometimes called the selectivity of the system. Assigning a measure to this selectivity is somewhat arbitrary. The simplest approach would be merely a graphical plot of amplitude versus frequency, i.e., the amplitude response. The steepness of the portions of the curve well removed from the passband is one kind of measure; it is sometimes called the skirt selectivity, since it involves the "skirts" of the response curve.

A better quantitative measure is provided when it is known that in the system where the amplifier will be used the undesired signals will be separated from the desired one by fixed frequency intervals. Thus, in broadcasting, the stations are assigned to channels separated by 10 kc, and the severest problem is to reject the signals in adjacent channels. The broadcast people have a term *adjacent-channel selectivity*, which is simply the ratio of the gain at midband to the gain 10 kc above or below midband. Similarly, *second-channel selectivity* would refer to 20 kc above or below midband. The same philosophy would apply to multichannel carrier systems.

In the general case, however, where frequency assignments are not so orderly, there is another numerical measure of the selectivity that has gained acceptance in recent years. This is the so-called *selectivity ratio or bandwidth ratio*. It is the ratio of the bandwidth at which the amplitude response is down 60 db from that at midband to the bandwidth at which the response is down 6 db. Thus it could be called the 60/6-db bandwidth ratio. Other levels could have been chosen, but these are convenient and serve the purpose of evaluating the response well outside of the desired passband. Typical values obtained are on the order of two for a good communications receiver,

and perhaps 12 for a radar system. Clearly, a value of unity would be the limiting case.

For the single-tuned amplifier, the selectivity ratio of a single stage is so large as to be almost useless, namely 577. As the number of stages is increased, however, the ratio decreases rapidly to a limiting value of 3.15. For six stages it has already dropped to 5.9. Other interstage arrangements will turn out to be superior, however.

$$\text{Selectivity Ratio for } n \text{ identical single-tuned stages} = \sqrt{\frac{\sqrt{10^6 - 1}}{\sqrt{4 - 1}}} \qquad (24\text{-}7')$$

24.7 Approximation: Commonly Used Functions for Approximating Constant Gain or Linear Phase

In the filter amplifier, the usual problem is to approximate an "ideal" amplitude response, such as in Figure 24-60, which cannot be physically realized with an amplifier containing only finite, lumped, linear networks.
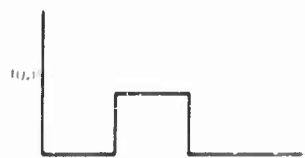
We want suitable gain functions which are physically realizable, and which we can later use to obtain element values for particular interstage networks (such as the single-tuned circuit of Section 24.6.1).

FIGURE 24-60

Notice that a cascade of identical single-tuned stages provides a crude approximation to constant gain. From Eq (24-65) the normalized amplitude response is:

$$\left(\frac{1}{\sqrt{1 + x^2}} \right)^n \qquad (24\text{-}65)$$

It really is not a very good function, either from the standpoint of the shape of the response curve being like Figure 24-60, or from the standpoint of conservation of gain-bandwidth. Two gain functions will be presented which are better in both respects, and can be realized almost as readily with the single-tuned or other interstage networks.

Another approximation problem is the obtaining of moderately effective filtering plus a good linear phase characteristic in the passband.

24.7.1 Maximally Flat Gain Function*

Also known as the "Butterworth" function (Reference 35), and as "ap-

*The term *maximal flatness* was first introduced in Reference 39, pages 347-362. See also Reference 40.

proximation in the Taylor sense," this function has for its normalized magnitude, or amplitude response, the following form:

$$\frac{1}{\sqrt{1 + x^{2n}}}$$

The shape of the response curve for various values of n (the number of stages) is shown in Figure 24-61. As n increases, the shape becomes more
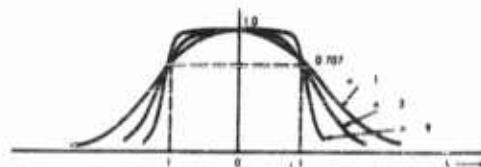
FIGURE 24-61

nearly the rectangle of Figure 24-60. The 3-db bandwidth remains constant. The function is always monotonic; i.e., decreases uniformly toward zero on either side of band center, and for a given n it represents "maximal flatness" in that the maximum number of derivatives $(2n-1)$ are zero at band center. This feature can be demonstrated as follows:

$$f(x) = (1 + x^{2n})^{-\frac{1}{2}} \tag{24-74}$$

Expand $f(x)$ in a power series:

$$f(x) = 1 - \frac{1}{2}x^{2n} + \frac{3}{4}\frac{x^{4n}}{2!} - \dots \tag{24-75}$$

Compare Eq (24-75) with the corresponding Taylor (McLaurin) series for $f(x)$: $\tag{24-76}$

$$f(x) = f(0) + f'(0)x + f''(0)\frac{x^2}{2!} + \dots + \frac{f^{2n-1}(0)x^{2n-1}}{(2n-1)!} + \frac{f^{2n}(0)x^{2n}}{(2n)!} + \dots$$

1.0	These terms missing in Eq (24-75); hence	This term and
	$f'(0) = 0$	higher even-order
	$f''(0) = 0$	terms are present
	$\cdots$	in Eq (24-75)
	$\cdots$	
	$f^{2n-1}(0) = 0$	

Thus the first $2n - 1$ derivatives are always zero at $x = 0$ for the function

$1/\sqrt{1 + x^{2n}}$. Remember that $x = 0$ corresponds to $f = f_0$, the frequency of maximum gain in the bandpass amplifier case.

The selectivity ratio of the maximally flat response is given below, which is seen to be similar to Eq (24-73) for cascaded identical single-tuned stages, but smaller for given n, and hence superior:

$$\text{Selectivity Ratio} = \sqrt[2n]{\frac{10^6 - 1}{4 - 1}} = \sqrt[n]{577} \qquad (24\text{-}77)$$

24.7.2 Equal-Ripple Gain Function

An alternative to the maximally flat approximation to constant gain in the passband is the equal-ripple response illustrated in Figure 24-62.
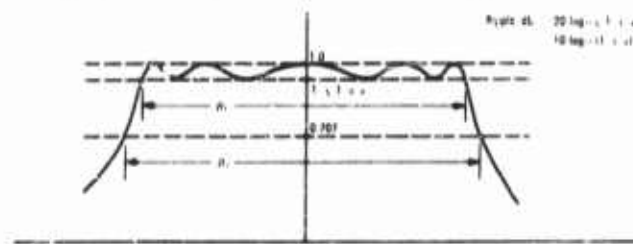
FIGURE 24-62

$$\text{Ripple, db} = 20 \log_{10} \sqrt{1 + \epsilon} \qquad (24\text{-}78)$$
$$= 10 \log_{10} (1 + \epsilon)$$

Such an amplitude response is readily devised. The magnitude of the ripples can be specified, although not their frequency spacing. This response function turns out to be advantageous from three standpoints: more gain can be achieved for the same bandwidth; the approximation to constant gain is usually better for steady-state applications than is the maximally flat response; and the selectivity ratio is better.

The analytic expression of the amplitude response of Figure 24-62 is given below. It contains Chebyshev polynomials, and because of this the approximation function is sometimes called the "Chebyshev response."

$$1/\sqrt{1 + \epsilon C_n^2 (x)} \qquad (24\text{-}79)$$

Where: ϵ = ripple parameter $= \left(\text{antilog}_{10} \dfrac{\text{Ripple, db}}{10} \right) - 1$

$$C_n(x) = \text{Chebyshev polynomials} \qquad (24\text{-}80)$$

$$C_1(x) = x \qquad\qquad C_4(x) = 8x^4 - 8x^2 + 1$$
$$C_2(x) = 2x^2 - 1 \qquad\qquad C_5(x) = 16\,x^5 - 20x^3 + 5x$$
$$C_3(x) = 4x^3 - 3x \qquad\qquad C_6(x) = 32x^6 - 48x^4 + 18x^2 - 1$$
$$C_{n+1}(x) = 2xC_n(x) - C_{n-1}(x) \qquad\qquad (24\text{-}81)$$

There is also a convenient expression for the general polynomial in terms of trigonometric functions:

$$C_n(x) = \begin{cases} \cos\,(n\,\cos^{-1}x) & \text{for } -1 < x < +1 \\[6pt] \cosh\,(n\,\cosh^{-1}x) & \text{for } |x| > 1 \end{cases} \qquad (24\text{-}82)$$

24.7.3 Linear Phase Response

In some systems either bandpass or low pass, it may be the objective to make the amplifier have, as closely as possible, a phase shift ϕ that varies linearly with frequency, or what is equivalent, a phase delay $d\omega$ that is constant with frequency. The amplitude response is taken as it comes.

The phase response is always an odd function of frequency with respect to band center, and may look as in Figure 24-63. The derivative on the other hand, will be an even function as in Figure 24-64. This suggests that one
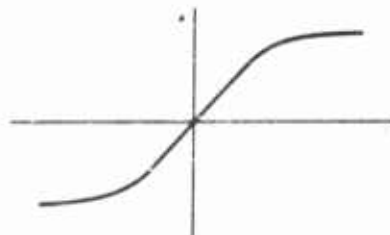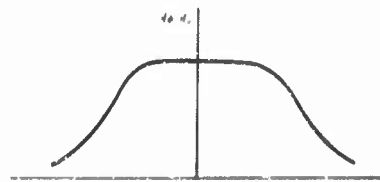
FIGURE 24-63   FIGURE 24-64

approximation to perfect linear phase (which is not physically realizable over an infinite frequency range) would be to make the phase delay maximally flat (References 41 and 42).

Equal-ripple approximations to linear phase (or constant delay) have received study, but no compilation of generally useful results has yet appeared.

24.7.4 Arbitrary Responses

The formal examples of amplitude or phase responses which have been presented are by no means the only useful ones. In some respects they are too circumscribed by theoretical limitations. For instance, it is not possible

to get both maximally flat amplitude and maximally flat delay in a simple amplifier structure; yet Bradley (Reference 43) has shown that one can devise a response which, from a practical standpoint, is very flat in both amplitude and delay and which has good selectivity as well; this is shown in Figure 24-65.
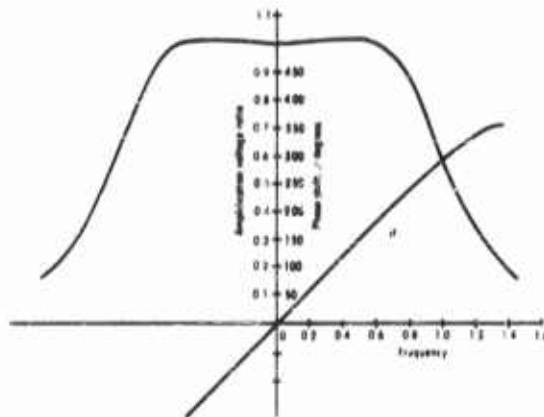
FIGURE 24-65

24.8 Stagger Tuning

The term stagger tuning refers to an amplifier comprising several stages in cascade, in which the stages are not tuned identically to the same frequency but are staggered at frequencies above and below the desired center frequency of the complete amplifier. Not only are the tunings of the individual stages nonidentical, but their bandwidths are also different.

The objectives of stagger tuning are twofold: (1) a greater gain-bandwidth is achieved than with a cascade of identical stages, and (2) a prescribed amplitude response, such as maximally flat or equal ripple, can be synthesized, either of which is preferable for filtering than is the response of identical stages.

Historically, the advantages and possibilities of stagger tuning were apparent to a few people several years before it became a widely used tech-..ique. The desirability of synthesizing a complicated gain function from simple networks in a multistage amplifier was first advocated by Butterworth in 1930 (Reference 35), although the gain-bandwidth advantage did not become apparent until Schieriemann's paper in 1939 (Reference 44). The latter paper was apparently not read by anyone in this country until about 1943, although Landon in the meantime had published a paper having to do with the maximally flat response function (Reference 39, pages 347-362 and

481-497). To Henry Wallman belongs the credit for "putting across" the stagger tuning technique (Reference 45),* in connection with wideband intermediate-frequency amplifiers used in the receiver of a radar system.

Wallman's work provided usable data for synthesizing the maximally flat amplitude response with single-tuned amplifier stages. This was extended by Baum (Reference 46) to include the equal-ripple function, and by Trautman, et al, to include other interstage networks (References 47, 48, 49, and 50).

The principal elements of the technique are to assign suitable values of center frequency f_0 and bandwidth or Q to each stage of a cascade of n single-tuned stages so that the desired amplitude (or phase) response is obtained. The derivation cannot be included here (Reference 51) but the results are easily presented and readily used.

24.8.1 Maximally Flat Amplitude Response

We shall follow Wallman's convention of distinguishing three cases, depending upon the relationship of bandwidth to center frequency. The first case is *narrow-band*, where the bandwidth is less than 5 percent of the center frequency. At the other extreme is the *wideband* case, where the bandwidth is 30 percent or more of center frequency. In between is what Wallman calls the asymptotic or *intermediate* case.

Tables 24-V, 24-VI, and 24-VII and Figures 24-66 and 24-67 use values from Valley and Wallman. The tables are given only through the staggered
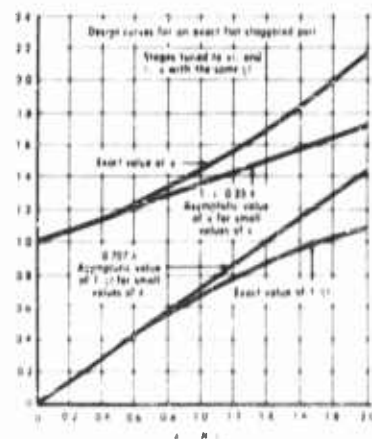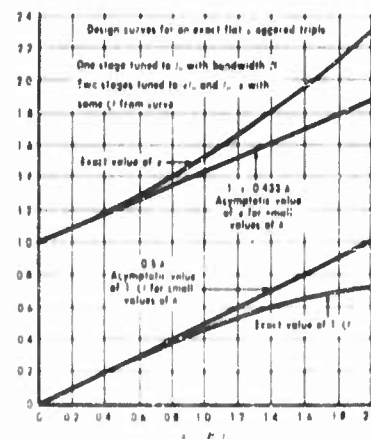
FIGURE 24-66   FIGURE 24-67

*The essential content of this report appears as Chapter 4 in Reference 2.

quadruple, although obviously they could be extended as far as desired. Practically, however, the pair and triples are the most widely used. Higher orders require stages with high Q, sometimes higher than can be obtained in the presence of loading due to tube input conductance. Also, the tune-up procedure is lengthened, since additional signal generator frequencies must be used. More about this later.

TABLE 24-V
NARROW-BAND STAGGER TUNING (MAXIMALLY FLAT)
$(B/f_0 < 0.05)$

1. Staggered Pair $(n = 2)$
   Two stages, tuned to $f_0 \pm 0.35B$, each having a bandwidth $0.707B$
2. Staggered Triple $(n = 3)$
   One stage tuned to f_0, with bandwidth B
   Two stages tuned to $f_1 \pm 0.43B$, with bandwidth $0.50B$
3. Staggered Quadruple $(n = 4)$
   Two stages tuned to $f_0 \pm 0.46B$, with bandwidth $0.38B$
   Two stages tuned to $f_0 \pm 0.19B$, with bandwidth $0.92B$

(Note: f_0 is the center frequency of the over-all amplifier, and B is the over-all 3-db bandwidth.)

TABLE 24-VI
WIDEBAND STAGGER TUNING (MAXIMALLY FLAT)

1. Staggered Pair $(n = 2)$
   Two stages tuned to $f_0 \alpha$ and f_0/α, having the same Q:

$$Q^2 = \frac{2}{4 + \delta^2 - \sqrt{16 + \delta^4}}$$

$$\left(\alpha - \frac{1}{\alpha} \right)^2 + \frac{1}{Q^2} = \delta^2$$

2. Staggered Triple $(n = 3)$
   One stage tuned to f_0 with bandwidth B
   Two stages tuned to $f_0 \alpha$ and f_0/α, same Q:

$$Q^2 = \frac{2}{4 + \delta^2 - \sqrt{16 + 4\delta^2 + \delta^4}}$$

$$\left(\alpha - \frac{1}{\alpha} \right)^2 + \frac{1}{Q^2} = \delta^2$$

(Note: f_0 is the center frequency (geometric center) of the over-all amplifier. B is the over-all 3-db bandwidth and $\delta = B/f_0$.)

TABLE 24-VII
INTERMEDIATE BANDWIDTH (MAXIMALLY FLAT)
$$\delta = B/f_0 \approx 0.05 \text{ to } 0.3$$

1. Staggered Pair ($n = 2$)
   Two stages tuned to $f_0\alpha$ and f_0/α; same Q:
   $$Q = 1.414/\delta \qquad \alpha = 1 + 0.35\delta$$

2. Staggered Triple ($n = 3$)
   One stage tuned to f_0 with bandwidth B
   Two stages tuned to $f_0\alpha$ and f_0/α; same Q:
   $$Q = 2.0/\delta \qquad \alpha = 1 + 0.433\delta$$

3. Staggered Quadruple ($n = 4$)
   Two stages tuned to $f_0\alpha_1$ and f_0/α_1; same Q_1:
   $$Q_1 = 2.63/\delta \qquad \alpha_1 = 1 + 0.46\delta$$
   Two stages tuned to $f_0\alpha_2$ and f_0/α_2; same Q_2:
   $$Q_2 = 1.088/\delta \qquad \alpha_2 = 1 + 0.19\delta$$

(Note: f_0 is the center frequency of the over-all amplifier, and B is the over-all 3-db bandwidth.)

The data in the table provide the necessary design data, and that would be the end of the story if electrical components always had the proper value of resistance, capacitance, etc. Such is not the case, of course, and particularly in wideband amplifiers where the principal capacitance is that due to the tube. There is substantial variation from one tube to the next, and hence each stage is usually made tunable over a sufficient range to allow for this. The inductance is readily tuned by means of a brass or powdered iron "slug", or core, moved into and out of the coil.

The tuning procedure is simplicity itself. A signal generator is connected to the input of the amplifier, and a vacuum-tube voltmeter to the output. If the amplifier is, say, a triple, with stages 1, 2 and 3, to be tuned to frequencies f_1, f_2 and f_3, according to Table 24-V, then the signal generator is first set to f_1, and Stage 1 tuned for a maximum voltmeter reading. Next, Stage 2 is adjusted with the signal generator set at f_2. Finally, Stage 3 is adjusted at frequency f_3. The order of adjusting the stages is completely unimportant.

Also, except for certain second-order effects, it is immaterial in which order the stages are connected.

The formulas of Table 24-VI are cumbersome to use, and hence should be avoided unless the bandwidth is really large. An *intermediate* region of B/f_0 can be defined, in which the calculation can be simplified, and yet retaining an accuracy of less than 1 percent. If B/f_0 is less than 0.3, the values of α and Q in Table 24-VI approach very closely an *asymptotic* limit; this is indeed called the "asymptotic case" by Wallman. For example, in the staggered

pair, α approaches $1 + 0.358$ and Q approaches 1.4148. These formulas for the intermediate case are given below in Table 24-VII. It will be noted from a comparison of the narrow-band and intermediate cases, Tables 24-V and 24-VII respectively, that the tuning of the low stages is the same, but not the high stages. In the narrow-band case the bandwidths of corresponding high and low stages are the same, whereas in the asymptotic case the Q's are the same. The differences, though, are quite small.

24.8.2 Equal-Ripple Amplitude Response

To obtain the equal-ripple response in the narrow-band case, it is only necessary to multiply the bandwidth of each stage by the factor $\tanh a$, where

$$a = (1/n)\ \cosh^{-1}\ (1/\sqrt{\epsilon}) \tag{24-83}$$

For the wideband and intermediate cases, the equal-ripple response can be obtained from the formulas found in Table 24-VIII.

<div align="center">

TABLE 24-VIII
WIDEBAND STAGGER TUNING (EQUAL RIPPLE)
Ripple factor $\tanh a$ as in Eq (24-83) and Figure 24-62.

$$\delta = \frac{B}{f_0} = \frac{\text{Desired over-all 3-db bandwidth}}{\text{Desired center frequency}}$$

</div>

1. Staggered Pair ($n = 2$)

   Two stages tuned to $f_0\alpha$ and f_0/α, having same Q:

$$Q^2 = \frac{2}{4 + R^2 - \sqrt{(R^2 - 4)^2 + 8\delta^2}}$$

$$\left(\alpha - \frac{1}{\alpha}\right)^2 + \frac{1}{Q^2} = R^2$$

$$R^2 = \frac{\delta^2}{2}\left(1 + \tanh^2 a\right)$$

2. Staggered Triple ($n = 3$)

   One stage tuned to f_0 with $Q = 1/(\delta \tanh a)$
   Two stages tuned to $f_0\alpha$ and f_0/α, same Q:

$$Q^2 = \frac{2}{4 + R^2 - \sqrt{(R^2 - 4)^2 + 12\delta^2}}$$

$$\left(\alpha - \frac{1}{\alpha}\right)^2 + \frac{1}{Q^2} = R^2$$

$$R^2 = \frac{\delta^2}{4}\ (3 + \tanh^2 a)$$

Note that the formulas of Table 24-VIII reduce to those of 24-VI (maximally flat) when the ripple is reduced to zero, making tanh $a = 1$. Thus 24-VIII includes not only equal ripple, but also maximally flat as a special case. Thus the gain-bandwidth factor diminishes as the number of stages increases.

24.8.3 Cascades of n-uples

A stagger-tuned cascade of n stages is called an n-uple, e.g., a quadruple for $n = 4$, but an n-uple where n is arbitrary.

Now, an n-uple can be designed according to the principles which have been set forth in the preceding pages for as many stages as desired. For practical reasons, however, the order n of the n-uples usually is not higher than three or four (occasionally as high as six). Hence, if more than three or four stages are required to obtain the required over-all gain, it is customary to cascade several triples (or quadruples). It turns out, of course, that when identical triples are cascaded, the bandwidth shrinks in a manner like that of cascaded single-tuned stages as given in Eq (24-68), except that the proper relationship now is:

$$\frac{\text{Bandwidth of } m \text{ staggered } n\text{-uples}}{\text{Bandwidth of 1 staggered } n\text{-uple}} = (2^{1/m} - 1)^{1/2n} \qquad (24\text{-}84)$$

Implicit in this equation is the fact that the higher the order n the slower will be the bandwidth shrinkage as n-uples are connected in cascade.

24.8.4 Gain-Bandwidth Factor

This useful figure of merit was defined in Section 24.6 and provides a means of comparing the amount of gain-bandwidth realizable from tubes having the same g_m/C when used with various interstages.

For one single-tuned stage, the gain-bandwidth factor was shown to be, by definition, equal to unity.

For a cascade of n identical single-tuned stages, the gain-bandwidth factor is:

$$\sqrt{2^{1/m} - 1} \qquad (24\text{-}85)$$

In marked contrast to this, the gain-bandwidth factor of n single-tuned stages arranged in a staggered n-uple (maximally flat) is always 1.00, regardless of n. Thus, stages can be added indefinitely without loss of gain-bandwidth factor.

When it becomes necessary to cascade staggered n-uples, the gain-bandwidth factor is given by Eq (24-84).

The numerical values which result from these formulas are both of interest and of practical importance. Table 24-IX shows the gain-bandwidth factors obtained with l tubes, from one to nine, used in various combinations.

TABLE 24-IX
GAIN-BANDWIDTH FACTORS

No. of Tubes l $(= m \cdot n)$	Identical Stages $\sqrt{2^{1/n}-1}$	Cascaded n-uples $(2^{1/m}-1)^{1/2n}$		
		m Pairs $\sqrt[4]{2^{1/m}-1}$	m Triples $\sqrt[6]{2^{1/m}-1}$	m Quadruples $\sqrt[8]{4^{1/m}-1}$
1	1.00			
2	0.64	1.00 $(m=1)$		
3	0.51		1.00 $(m=1)$	
4	0.44	0.80 $(m=2)$		1.00 $(m=1)$
5	0.39			
6	0.35	0.71 $(m=3)$	0.86 $(m=2)$	
7	0.32			
8	0.30	0.66 $(m=4)$		0.90
9	0.28		0.80 $(m=3)$	

To use the table, refer back to Eq (24-69) where the gain-bandwidth factor was defined. As an example, suppose one wishes to compare nine tubes used as identical stages or as three staggered triples. The ratio of over-all bandwidth obtainable for the same over-all gain is:

$$\left[\frac{0.80}{0.28}\right] = 2.86$$

On the other hand, the ratio of gain obtainable for the same over-all bandwidth is:

$$\left[\frac{0.80}{0.28}\right]^9 = 1.3 \times 10^4$$

Comparison of the gains for equal bandwidths emphasizes the fact that an amplifier made up of a large number of identical stages is indeed an inefficient device. The contrast for the case of equal gain is not so startling, but the same phenomenon is at work.

24.8.5 Practical Design Information

The attempt in this chapter has been to provide the underlying theory and

some physical intuition in the matter of how stagger tuning works. Active workers in a field such as this inevitably produce helpful graphs, tables, nomograms, and the like, to shorten the time required for numerical designs (References 52, 53, and 54). The basic work of Wallman is usually quite adequate for the maximally flat case, in particular, Chapters 4 and 8 in Reference 2 contain many practical details. For the equal-ripple case, the papers by Baum and by Wittenberg may prove useful (References 46 and 52).

24.8.6 Design Example

It is instructive to carry through an example, especially to show how certain graphical aids can be devised and put to use.

Suppose that in a given system there is needed a bandpass amplifier to provide a gain (voltage amplification) of 60 db with a bandwidth of 7 mc. The center frequency is of no consequence in the initial phases of the design procedure, and in fact is needed only when one comes to calculate the interstage inductances. System considerations, such as reducing the number of tube types, limit the available tubes to the 6AK5, 6CB6 and 6AH6. Measurements on the wiring situation in which the tubes will be mounted show that a total interstage wiring capacitance of $5\mu\mu f$ can be expected.

The design questions which we shall answer here are: (a) Which tube should be used? (b) Which combination of single-tuned stages will meet the gain-bandwidth requirements? (c) Which will require the fewest tubes? (d) Which will give the best selectivity ratio?

(a) Choice of tube:

Shown in Figure 24-68 is a plot of Eq (24-60) for several currently used tubes, including the three allowable types for the example at hand. The points on the graph do not include wiring capacitance, however, so each of our three must be translated upwards by $5\mu\mu f$. Because of the logarithmic capacitance scale, the 6AH6 is displaced the least, hence ends up as the best tube choice. The gain-bandwidth for the 6AH6 with this wiring capacitance is 84 mc.

(b) Choice of stagger combinations:

The gain-bandwidth performance can be displayed conveniently in the graphical presentation devised by Wightman (Reference 54), which is represented here as Figure 24-69. The "normalized bandwidth" is obtained by taking the actual over-all bandwidth required of the amplifier (7 mc in our example) and dividing it by the $g_m/2\pi C$ of the tube (84 mc). Thus, for the example at hand, the normalized bandwidth is 7/84 or about 0.083. Looking at Figure 24-69 it will be seen that all curves which cross the vertical line
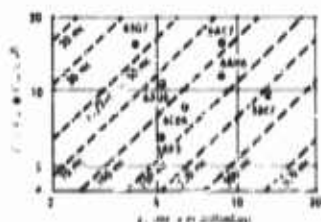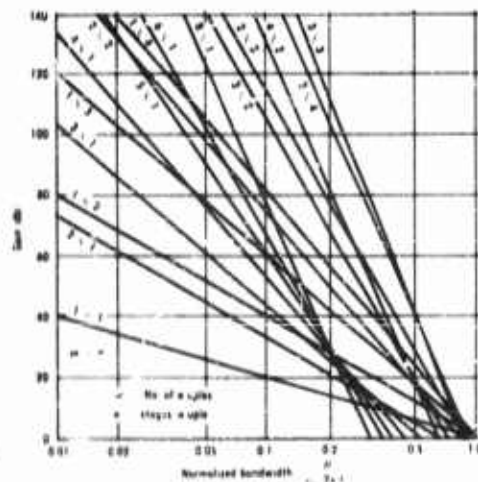
F. 24-68   FIGURE 24-69

through 0.083 at a level of 60 db or greater will meet the requirements. The lowest of these are:

> 1 × 3 (one staggered triple; 3 tubes)
> 2 × 2 (two staggered pairs; 4 tubes)
> 1 × 4 (one staggered quadruple; 4 tubes)

(c) Fewest tubes:

It is evident that the staggered triple requires the fewest tubes. This may not be the practical answer, however. System requirements may favor the better selectivity ratio of the 2 × 2 or 1 × 4 in spite of the extra tube (see below). Also practical considerations of variability of tube characteristics may call for a margin of safety, thus favoring the other combinations.

(d) Selectivity Ratio:

The selectivity ratio for a cascade of identical stages has already been given in Eq (24-73), and for the maximally flat function corresponding to a single n-uple in Eq (24-77). For a cascade of m identical n-uples, the selectivity ratio is:

$$\text{Selectivity Ratio} = \sqrt[2n]{\frac{\sqrt[m]{10^n} - 1}{\sqrt[m]{4} - 1}} \qquad (24\text{-}86)$$

It is instructive to display this relationship graphically as in Figure 24-70, which permits one to see quickly the relative merits of various combinations.

For the example at hand, the selectivity ratios for the three alternatives are:

$m \times n$	Selectivity ratio
1×3	8.4
2×2	5.7
1×4	4.9

System requirements will have to govern the choice here. It would appear that in going to four tubes, one might as well use the 1×4 combination and achieve the better selectivity ratio and higher gain. Here again, however, the system situation might favor the 2×2, since it has only two different inter-stage types to manufacture and align.
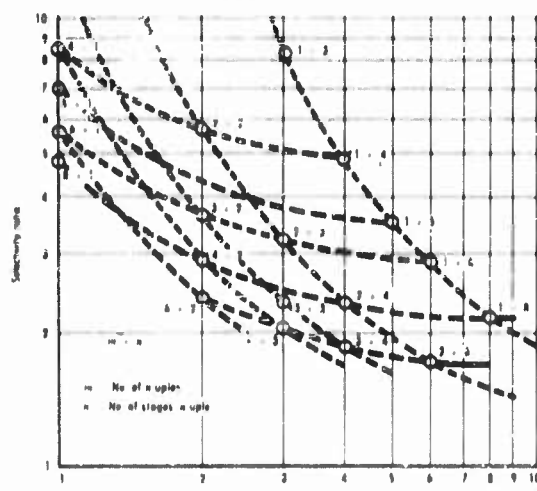
FIGURE 24-70

24.9 The Double-Tuned Interstage: A Four-Terminal Network

An alternative means of realizing amplifier gain functions, in contrast to the single-tuned interstage network employed thus far, will be presented in this section.

Nothing will be added to the approximation problem. Maximally flat and equal-ripple responses are still our most useful approximations to constant gain in the passband. We know what sort of complex gain functions will produce these responses; i.e., poles on a circle or on an ellipse, respectively. The task then is to find what sort of pole-zero arrangement comes out of the double-tuned circuit, and how the pole co-ordinates are related to the circuit parameters.

The double-tuned circuit, sometimes called transformer coupling, is a logical extension from the single-tuned circuit. It represents the general process of adding more circuit complexity in exchange for improved perfor-

mance. The improvement is of two kinds: a better gain-bandwidth factor and a better selectivity (better in that one stage provides a two-pole response, instead of a one-pole as with the single-tuned stage).

There are several equivalent forms of the double-tuned circuit, illustrated in Figure 24-71, the most common being the inductive coupling and the pi equivalent.
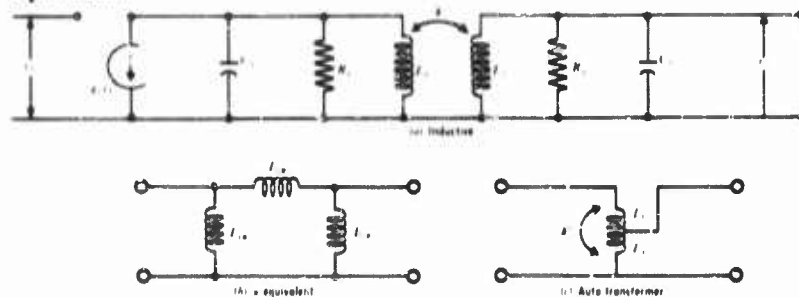
FIGURE 24-71

Initially we shall confine the discussion to the inductively coupled arrangement, specified by the primary inductance L_1, the secondary inductance L_2, and a coupling coefficient k, defined in the conventional way as:

$$k = M/\sqrt{L_1 L_2} \qquad (24\text{-}87)$$

where M is the mutual inductance between the two coils L_1 and L_2.

We can further define a primary resonant frequency ω_1 and a primary Q_1, and likewise for the secondary:

$$\omega_1 \triangleq 1/\sqrt{L_1 C_1} \qquad\qquad \omega_2 \triangleq 1/\sqrt{L_2 C_2} \qquad (24\text{-}88)$$

$$Q_1 \triangleq R_1/\omega_1 L_1 \qquad\qquad Q_2 \triangleq R_2/\omega_2 L_2 \qquad (24\text{-}89)$$

There are several cases of practical importance, which we subdivided first into the narrow-band and wideband situations. Let us consider the narrow-band case first, for which $k \ll 1$.

Next we consider two cases: (1) equal Q; i.e., $Q_1 = Q_2$, and (2) one Q infinite; e.g., $Q_1 = \infty$.

Maximally Flat Amplitude Response—Equal Q's.

The relationships in (24-90) or (24-91) are definitive for maximal flatness, and the ones usually found in the handbooks for the proper adjustment of a double-tuned circuit. The coupling coefficient k, so defined, is sometimes

known as transitional or critical coupling, since it is the cross-over value between a single-peaked and a double-peaked amplitude response.

$$k = \frac{1}{\sqrt{2}} \frac{B_r}{\omega_0} = \frac{1}{\sqrt{2}} \frac{B}{f_0} \qquad (24\text{-}90)$$

where B_r is in radians per second and B is in cycles per second.

$$\text{Critical coupling, } k_c \triangleq \frac{1}{\sqrt{Q_1 Q_2}} = \frac{1}{Q} \qquad (24\text{-}91)$$

Equal-Ripple Response

The equal-ripple response could be similarly realized by increasing the Q by multiplying Q by $1/(\tanh a)$.

Gain Bandwidth Factor

Now that it has been shown that a desired amplitude response can be realized, what about the gain-bandwidth factor? If one takes the value of k defined by (24-90) and uses it to solve for the gain at band center and multiplies this by the bandwidth B (cps), the following obtains:

$$|A(j\omega_0)| B = \left[\frac{g_m}{2\pi(2\sqrt{C_1 C_2})} \right] \sqrt{2} \qquad (24\text{-}92)$$

This is the gain-bandwidth product, and it contains a term dependent only upon the tube; namely $g_m/2\pi(2\sqrt{C_1 C_2})$. This quantity corresponds to $g_m/2\pi(C_1 + C_2)$ which was the gain-bandwidth product of the single-tuned stage, and is very closely the same unless C_1 is very different from C_2 (the difference is 6 percent for a 2:1 ratio). The new form of this factor is indeed the gain-bandwidth product for the single-tuned circuit if an ideal transformer is included. The important fact is that the double-tuned circuit, with equal primary and secondary Q is better by $\sqrt{2}$, and hence the gain-bandwidth *factor* of the circuit is $\sqrt{2}$.

Unequal Q's

The next case of interest is that in which one of the Q's is infinite. This condition can only be approximated in practice, since the primary is always loaded by the plate resistance of one tube, and the secondary by the input conductance of the other. Both primary and secondary have inevitable circuit losses. Nevertheless, in wideband applications the required Q is so low that,

In contrast, the minimum Q of either primary or secondary (primary in particular, at high frequencies) will be so large in proportion that the results predicted by assuming this Q to be infinite are approximated quite closely.

The gain-bandwidth factor for this case can be found in the same way as for the equal-Q case. The result is that the gain-bandwidth factor when one Q is infinite is 2.0, instead of $\sqrt{2}$ for equal Q's.

Cascading of Stages

When maximally flat double-tuned stages are used in cascade, the bandwidth narrows. The narrowing factor is the same for any two-pole stages, and hence must be the same as for staggered pairs, namely the factor obtained from Eq (24-84) for $n = 2$:

$$\frac{\text{Bandwidth of } m \text{ stages}}{\text{Bandwidth of 1 stage}} = \sqrt[4]{2^{1/m} - 1} \tag{24-93}$$

The gain-bandwidth factor for m stages is the value of (24-93) multiplied by either $\sqrt{2}$ or 2, depending on whether $Q_1 = Q_2$ or $Q_1 = \infty$.

Stagger Damping

The term "stagger damping" was coined, apparently by Wallman, to describe a form of stagger tuning using double-tuned circuits. Only the narrow-band situation is permitted. The technique permits the synthesis of, for example, a four-pole maximally flat response with two double-tuned stages in cascade. The stages are not identical, but are "staggered", except that both stages are tuned to the same center frequency and it is the Q, or damping, that differs between the two stages.
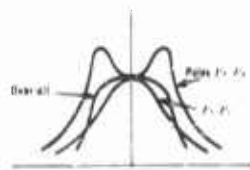
The basic principle of stagger damping can be demonstrated readily by means of Figure 24-72. The first stage will yield a single-peaked response and the second a double-peaked response. The two combine to give a maximally flat over-all response.

FIGURE 24-72

Both stages have the same center frequency f_0, which is the desired center frequency of the pair. With an equal Q design, the k and Q are as follows:

$$
\begin{aligned}
\text{Stage 1} \quad & k = (B/f_0) \cos (\pi/8) \\
& Q = 1/|(B/f_0) \sin (\pi/8)| \\
\text{Stage 2} \quad & k = (B/f_0) \cos (3\pi/8) \\
& Q = 1/|(B/f_0) \sin (\pi/8)|
\end{aligned}
\tag{24-94}
$$

For unequal Q, specifically for $Q_1 = \infty$, both stages have the same coupling coefficient $k = B/f_0$. The values of Q are as follows:

$$\text{Stage 1} \qquad Q_3 = 1/[2(B/f_0) \sin (\pi/8)]$$
$$\text{Stage 2} \qquad Q_4 = 1/[2(B/f_0) \sin (3\pi/8)] \qquad (24\text{-}95)$$

The technique is obviously not limited to the four-pole case. Any even number of poles can be used, each pair corresponding to one double-tuned stage. The practical limits are excessive Q's and extreme precision of adjustment for the higher order cases.

The Wideband Case

In the wideband case the center frequencies of the stages are necessarily nonidentical, thus introducing additional complexity. The rewards are substantial, however, in gain-bandwidth performance compared to stagger tuning with single-tuned circuits (Reference 55). For example, a staggered pair will give 12 db more gain for the same bandwidth than will a pair of staggered single-tuned stages. A triple will give 18 db more.

A response curve obtained with a staggered pair is shown in Figure 24-73.
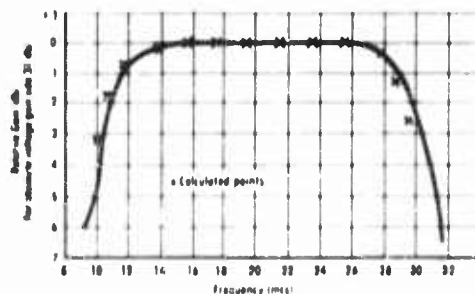
FIGURE 24-73

Notice that the curve has approximate arithmetic symmetry, rather than geometric, relative to the center frequency (20mc.) In contrast, a staggered pair of single-tuned stages would produce geometric symmetry, i.e., a steeper fall off on the low-frequency side.

Equal-Ripple Response

It is also possible to provide an equal-ripple response with double-tuned circuits. For a single stage and for bandwidths which are not excessively large, the data published by Dishal (Reference 36) can conveniently be used. For extremely wide bands, and for stagger-tuned stages, it is theoretic-

a''y possible to provide exact equal-ripple response (Reference 49), but
v   ble design data are not yet available.

The Capacitance-Coupled Circuit

The discussion up to now has been confined to the situation where the
primary and secondary circuits were inductively coupled together, either with
mutual inductance or with the pi or T equivalent (Figure 24-71). There is
an alternative case, which is of both theoretical and practical interest. This
case is called capacitive (or capacitance) coupling, and is illustrated in
Figure 24-74. There is no coupling from primary to secondary except
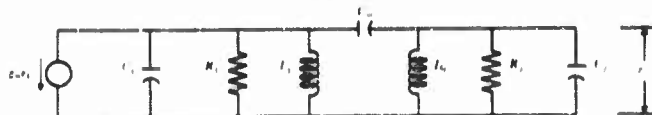
FIGURE 24-74

through the capacitance C_m. A coupling coefficient can be defined for the
network, analogous to the inductively coupled circuit:

$$k \triangleq \frac{C_m}{\sqrt{C_1 C_2}} \tag{24-96}$$

In a narrow-band situation, i.e., small B/f_0, capacitance coupling and in-
ductive coupling are about equally useful. As the bandwidth increases, how-
ever, certain disadvantages appear for the capacitive case. The shape of the
amplitude response curve is highly unsymmetrical, dropping off quite
sharply on the low-frequency side as depicted in Figure 24-75. This dissym-
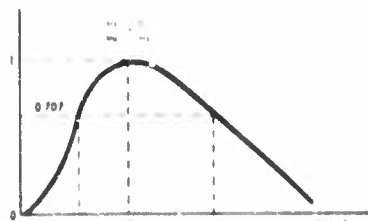
FIGURE 24-75
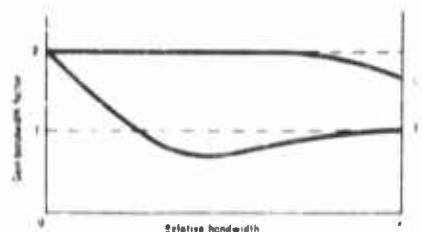
FIGURE 24-76

metry is substantially greater than with the single-tuned circuit, where as
described earlier there is exact geometric symmetry.

Moreover, the gain-bandwidth factor of the capacitance-coupled circuit
is highly unfavorable in wideband situations. For comparison with the in-

ductively coupled case, there is plotted in Figure 24-76 the curves of gain-bandwidth factor for $Q_1 = \infty$.

The Autotransformer

Another form of double-tuned interstage, which is really a special case of inductive coupling, is the autotransformer shown in Figure 24-71(c). While this scheme is well known at 60 cps, it has not received wide employment at radio frequencies. It nevertheless is quite applicable, and the design can be straightforward (Reference 56). It has particular advantage in wideband double-tuned amplifier stages with large bandwidth and large ratio of C_1 to C_2; in fact, it turns out conveniently that the autotransformer is physically realizable in those regions of operating conditions where the pi equivalent is not (because of one or more negative elements) as seen in Reference 55.

Selectivity Ratio

The selectivity ratio of one maximally flat, double-tuned stage is the same as that of the maximally flat staggered pair. Similarly, the selectivity ratio of cascades of identical stages is the same as for cascades of pairs. Hence, both Eq (24-86) and Figure 24-70 apply, provided one takes values only for $n = 2$.

Stagger damping or stagger tuning in the wideband case can also be studied from Figure 24-70. For a maximally flat pair of double-tuned stages take $n = 4$; for a maximally flat triple, take $n = 6$, etc.

24.10 Low Noise Traveling-Wave Tubes

The helix type traveling-wave tube, which is described in more detail in Chapter 26, is valuable in ferret and ECM applications, because it is able to amplify over extremely wide frequency bands. It is convenient and customary to cover octave bandwidths up to eight kmc. At higher frequencies it is possible to cover still greater absolute bandwidths, but in ratios progressively decreasing from an octave. Thus eight to twelve and twelve to eighteen kmc are typical.

In most situations it is important that the waveform of the received signal be accurately reproduced. Therefore phase and amplitude distortion are undesirable. Moreover, if some or all of the signals are weak, it is necessary that the first amplifier have a *low noise figure*. Otherwise, the signal will be obscured by random disturbances called noise.

In traveling-wave tubes the principal source of noise is randomness in the basic electron stream, which produces amplification by interacting with the electromagnetic fields within the helix. Accordingly, the tube becomes progressively "quieter" as this source of noise is suppressed to a negligible value by making the electron flow more smooth and orderly. Additional noise can

be introduced by thermal agitation of electrons within the metal of the helix itself. However, this effect, which depends upon the temperature and attenuation constant of the input portion of the helix, is ordinarily small and when necessary can be further reduced by cooling or by use of a low-loss material.

The principal constructional features of a low-noise traveling-wave tube are shown in Figure 24-77. Design interest centers on the cathode and as-
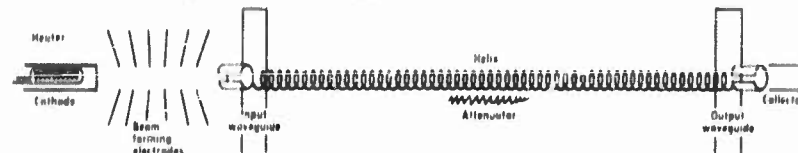
FIGURE 24-77.   Low-noise, helix-type traveling-wave tube.

sociated electrodes which make up the electron gun that produces the electron beam. Because the behavior of such guns is extremely complex, a large amount of work has been required to reduce noise levels to their present value. The lowest practical cathode temperature is used, the geometry is carefully controlled so that electrons move in smooth non-crossing paths, and their velocity is increased gradually by use of about half a dozen separate ring-shaped electrodes, each at an increased potential.

The electron beam is prevented from spreading by a uniform magnetic field which is accurately parallel to the axis of the helix. After passage through the helix the electrons are captured in a small metal cup or collector. To avoid undesirable gain variations and increase of noise it is important to arrange matters so that no appreciable number of electrons released by secondary emission from the collector are able to traverse the helix in the reverse direction. Also, it is necessary to achieve and preserve a very high degree of vacuum. Otherwise the ions formed by bombardment of the residual gas perturb the electron beam and add noise to the output signal by producing random modulation of the phase and amplitude.

Typical low noise traveling-wave tubes have saturation power output levels of about one milliwatt, values of gain in the neighborhood of 25 db and dynamic range near 50 db.

Noise figures typical of contemporary production-type traveling-wave tubes are shown in Figure 24-78. These values are strictly comparable with noise figures typical of triodes and parametric amplifiers. It must be understood that such a chart is necessarily general, and that somewhat lower as well as considerably higher noise figures will be met in individual situations. Neither triodes nor parametric amplifiers are capable of matching traveling-wave tubes in bandwidth, though the latter are currently subject to very rapid improvement (References 57 and 58).
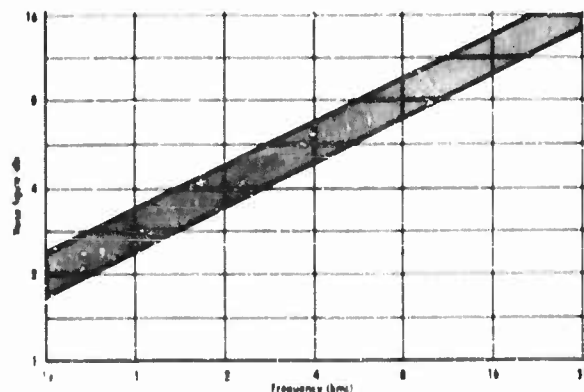
FIGURE 24-78. Typical values of noise figure.

24.11 Parametric Amplifiers

The term *parametric amplifier* is a shortening of *amplifier based on time-variable parameters*. Unlike conventional amplifiers, the power increase comes from an alternating rather than direct current source. Parametric amplification *can* be achieved in a linear system in which an inductance or capacitance is varied periodically by mechanical means. However, electronic devices are capable of frequencies far higher than anything which can be achieved mechanically. Therefore, such linear systems are of relatively little importance.

Practical parametric amplifiers employ nonlinear elements, in which the inductance or capacitance presented to the weak signal is periodically varied by a much larger *pump* signal which is usually at a substantially higher frequency. Both ferromagnetic and ferroelectric materials are characterized by nonlinear behavior such that the incremental permeability or permittivity is subject to wide variation during the cycle of a sufficiently strong (pump) signal. Also, the effective capacitance of a semiconducting diode is quite sensitive to the total back bias applied.

Perhaps the simplest example of parametric excitation is provided by a pendulum in which the length is periodically varied. Such a system results if the line to a plumb-bob is passed through an eyelet and pulled at appropriate times. Referring to Figure 24-79, it is easy to see that energy will be delivered to the bob and that the amplitude of the oscillation will tend to increase if the line is pulled during those intervals when the bob is approaching the center of its path and is released so that the pendulum lengthens when the motion is away from center. It is seen that the line is pulled *twice* while the pendulum describes one full cycle of oscillation. Therefore it follows that the frequency of the driven or generated oscillation is exactly half
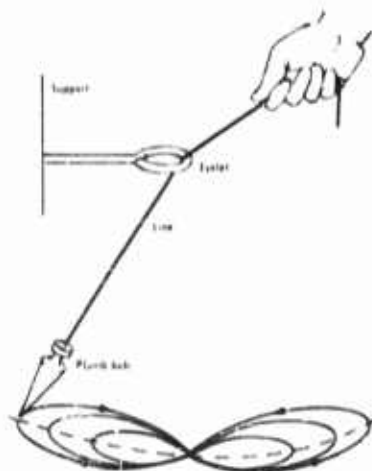
Figure 24-79.   Variable-length pendulum.

that of the driving or pumping signal. This two-to-one relationship is useful in frequency dividers but represents only one special case of parametric excitation.

By analogy with the variable-parameter pendulum we are led to the correct conclusion that an electric circuit will oscillate if its tuning is periodically changed by variation of either the inductance or capacitance. In fact, the ordinary induction alternator is an example of this relationship. Here again, the simplest situation is a singly-resonant circuit in which the value of the inductance or capacitance varies at twice the natural frequency.

A more complicated but much more useful situation arises when the electric circuit has two natural frequencies and the pumping signal has a frequency which is the sum of these two. Simultaneous (and closely coupled) oscillations arise at the two natural frequencies of such a circuit provided the losses are low enough in relation to the power level of the pumping signal. Oscillations cease but a substantial negative resistance remains at both frequencies if the level of the pumping signal is reduced somewhat below that which causes oscillation. Such negative resistance is the basis of most practical parametric amplifiers. In such a doubly-resonant system a relatively weak input signal at one of the frequencies produces a considerably larger power output at either frequency. Therefore, the two-frequency parametric amplifier is also capable of operating as a frequency changer.

The principal advantage of parametric amplification is that it adds very little noise to the input signal. This desirable property stems from the fact that thermal noise is associated with the resistive rather than the reactive

portion of circuit impedance. Therefore, a purely reactive circuit would con-
tribute no noise whatever, regardless of its temperature. This ideal situation
is, of course, never realized in practice. However, it is approximated well
enough so that under favorable circumstances a parametric amplifier oper-
ating at room temperature is capable of contributing no more noise than
that due to a pure resistance at 80° K. That is, in terms of a source at room
temperature the noise figure is well below one decibel. More representative
values of the noise figure to be expected of a parametric amplifier are given
in Figure 24-78.

A typical microwave parametric amplifier uses a back-biased semiconduct-
ing diode in a cavity resonator system tuned to two frequencies; the input or
signal frequency, and an auxiliary or *idler* frequency which is usually con-
siderably higher. In addition, the system must freely transmit the pump
frequency which is the sum of the two.
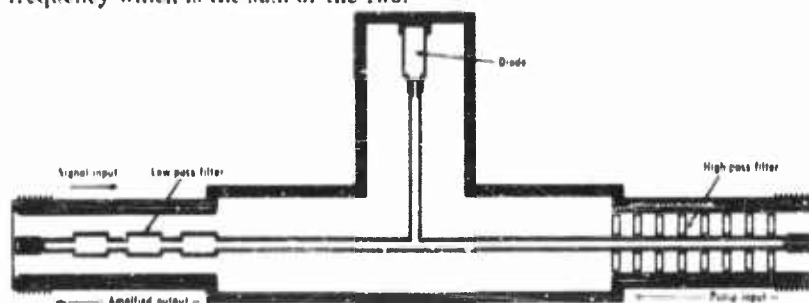
Figure 24-80.   Parametric amplifier.

A system having suitable characteristics is shown in Figure 24-80. The
lengths and impedances are such that nearly all the power supplied at the
pump frequency is delivered to the crystal diode and that virtually none
leaks through to the signal source. The coaxial structure also has a low-loss
resonance at the signal frequency and another at the idler frequency. The
junction between the input low-pass filter and the resonant structure is
chosen so as to provide a good power transfer to the crystal without seriously
loading the signal-frequency resonance.

A system of this sort readily produces a gain of about 20 db with a band-
width of 2 to 10 mc in the frequency range 200 to 3000 mc. The pump
delivers a power of about 50 milliwatts at a frequency about four times the
signal frequency. The saturation signal level is about — 30 dbm, the dynamic
range about 80 db and the noise figure one to four db.

The noise figure tends to improve as the pump frequency is increased, but
this tendency is offset by the fact that available pump signals become noisier

and less stable with increasing frequency. Therefore, in any given situation there is an optimum pump frequency.

It is appropriate to close this section with a brief note concerning the solid-state maser, which achieves much the same result in a quite different way. In the maser the awkward problem of controlling three natural frequencies in a cavity resonator is avoided by selecting a material in which appropriate resonances are inherent in the structure of the substance itself. This is possible because a number of materials such as ruby have the property of paramagnetic resonance, which may be varied by means of an external magnetic field. In the parametric amplifier the power source is a pump at the highest frequency, and amplification or frequency conversion may be obtained at either of the two lower frequencies.

A good general account of parametric amplifiers is given in Reference 59. A more detailed treatment with comprehensive bibliography appears in Reference 60, and additional theory is given in Reference 61.

Section II: Filters

24.12 Filters

A *filter* is a structure or system which is capable of selecting or discriminating between signals having different characteristics. The term originated in connection with frequency-selective filters consisting of combinations of passive linear low-loss reactive elements. Low-pass and bandpass filters are familiar examples of such *frequency filters*. The wide use of pulse circuits and techniques made it clear that signals are sometimes more readily identified by their time pattern than by their frequency spectrum and demonstrated the need for another class of filters which discriminate between signals in terms of direct waveforms, and are referred to as *time-domain* or *correlation filters*.

Because unique relationships, expressible in terms of Fourier and Laplace transforms, exist between the waveform (time domain) and spectrum (frequency domain) specifications of a given signal, it is clear that frequency and correlation filters are not independent. However, the relationships in question are relatively complex. Therefore, in typical situations one of these two alternate approaches to the filter problem is far simpler and more straightforward than the other.

24.12.1 Frequency Filters

A frequency filter ordinarily consists of a combination of low-loss inductive and capacitive elements. The most general response characteristics and the
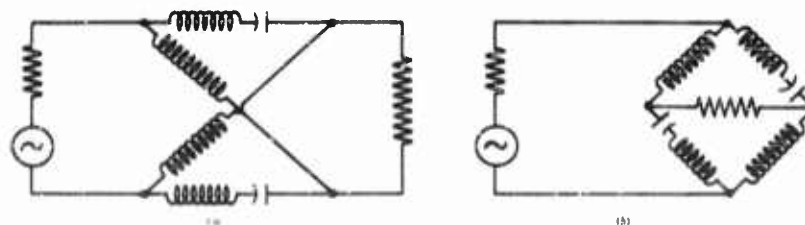
FIGURE 24-81. A simple low-pass filter. (a) Lattice representation. (b) Bridge representation.

most powerful design methods are associated with the lattice or bridge structure illustrated in Figure 24-81. Such filters are characterized by two different pairs of identical reactive arms, a useful resistive load, and a source which is also a pure resistance. Ordinarily, but not always, the source and load resistances are equal. The design ordinarily neglects the energy losses which result from unavoidable resistance in the reactive elements, and it is always desirable to minimize such resistive losses.

The lattice arrangement is unattractive in practice for several reasons. The number of elements required is large, neither load nor source may be grounded, and it is quite difficult to adjust the various elements to the precision required. Practical filters are ordinarily constructed in the ladder configuration shown in Figure 24-82. With few exceptions the unbalanced ar-
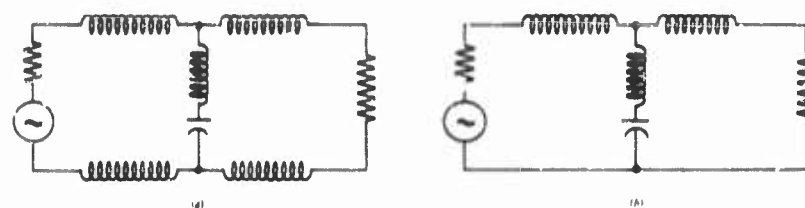
FIGURE 24-82. Ladder-type filters equivalent to Figure 24-81. (a) Balanced or symmetrical form. (b) Unbalanced or common-ground form.

rangement is preferred because it permits a common ground connection and uses a minimum number of elements. A further advantage of this configuration is that it is easy to compound the discrimination or frequency selectivity of several different filter sections by connecting them in tandem or cascade to form a composite filter. The prototype and m-derived filters of Zobel and Shea (Reference 62) are illustrative of this design technique.

The performance of frequency filters is ordinarily specified and measured in terms of a single-frequency sinusoid wave, variable as required. In typical situations it is desirable to transmit without attenuation all signals having

frequencies within a certain range called the passband and to reject signals at all other frequencies. It is important to note that conventional filters reflect toward the source all the energy which is not transmitted to the load. Therefore, they are characterized by large reflection coefficients outside the passband. Inevitably there are certain frequencies at which partial transmission occurs, but it is usually possible and desirable to limit these effects to relatively narrow transition or guard bands.

While filter requirements can b   ⁓ fied in terms of words or tabulated numbers, it is usually preferable to .   graphical representation such as that shown in Figure 24-83. It is seen that no requirement exists in the frequency
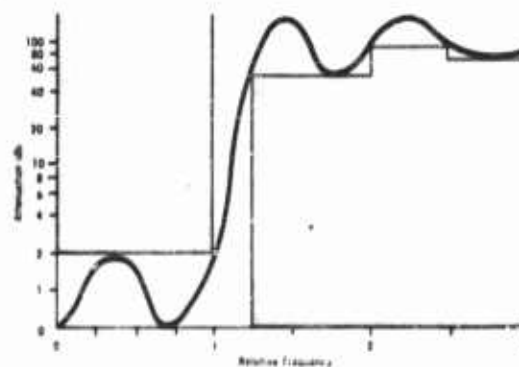
FIGURE 24-83.   Graphical specification of low-pass filter characteristic

interval 1.00 to 1.25. In the passband the loss must not exceed some small value such as 2 db. In the rejection band above 1.25 the loss must equal or exceed the various values indicated. Because the response of physical filters is always continuous, the most economical filter meeting the given specifications has a characteristic such as that shown in the heavy line.

At low frequencies it is possible and desirable to construct appropriate frequency selective filters of lumped inductive and capacitive elements. However, these elements become unattractive at frequencies above about 100 megacycles because the losses are excessive, the power handling capability is inadequate, and the stability is poor.

At frequencies ranging from about 100 to 1000 megacycles it is possible to construct excellent filters from sections of coaxial or (occasionally parallel-wire) transmission lines. Transmission lines are characterized by the fact that their inductive and capacitive reactances are smoothly distributed along their length. Therefore it is impossible to design such filters by the methods developed for lumped elements. Fortunately, the resonant properties of quar-

ter- and half-wave transmission line sections are quite similar to those of LC lumped circuits. Therefore, excellent approximate designs may be had by simple extensions of the methods used to design lumped-circuit filters.

At frequencies upwards of about 1000 mc it is necessary to abandon transmission lines and employ resonant cavities. Such cavities are mechanically stable and have extremely low intrinsic losses and therefore permit the construction of very selective filters. Unfortunately, such cavities are capable of resonating in a variety of ways. Each of these resonances is identified with a particular frequency and pattern of electric and magnetic fields and is referred to as a *mode*. To avoid the creation of unwanted (spurious) pass or rejection bands it is necessary to make the dimensions of the various cavities dissimilar in such a way that their unwanted responses do not coincide.

24.12.2 Correlation Filters

Correlation filters differ from frequency filters in several important ways. They often employ active elements such as vacuum tubes, the behavior is usually nonlinear, and time delay networks are likely to be included. Such filters are most conveniently specified in terms of waveforms in the time domain and tend to be more specialized than frequency filters. The analysis and design of time-domain filters have received considerably less study than has been devoted to frequency filters. Because the behavior of such filters is more complicated and capable of much wider variation than frequency filters, it follows that the specification of such filters is still very incomplete.

The general features of a correlation filter are illustrated by the range gate which has been used in many radar systems. The purpose of the range gate is to accept all signals which lie within a specified interval of time and to reject signals which occur at other times. The essential elements of such

a gate or filter are shown in Figure 24-84. A basic time reference is established by the clock or pulse repetition frequency generator. Signals, in this case a train of uniformly spaced short pulses, are produced by the signal source. The returned signals, in addition to being delayed, are ordinarily greatly attenuated and considerably contaminated with noise. They are amplified

Figure 24-84. System with range gate.

and converted to a convenient frequency before delivery to the electronic switch or gate, which is central to the present discussion. The position of the gate is controlled by an actuating pulse which is appropriately delayed to correspond to the signal of interest. In the present situation it is clear that optimum results are obtained if the gate is actuated for exactly

The other sources of radiation in this aircraft will be the engine intakes which may have temperatures as high as 300°C and the hot metal parts of the engines. The total output of the six engines will probably not be more than 4 kilowatts per steradian in the 4.3 micron region. The view of the hot metal and gases within the engine will be somewhat restricted by long exhaust nozzles.

21.7 Radiation Characteristics of Air-to-Air Rockets

The infrared radiation characteristics of rockets are also of interest in a discussion of infrared countermeasures. This is because use can be made of the radiation from the rocket motors of missiles in warning systems. (See section on engagement warning systems, Section 22-10.) Ultraviolet measurements on rockets have also been made in this connection.

Table 21-IV shows radiation characteristics of air-to-air rockets in the infrared in kilowatts per steradian at nose-on aspect (Reference 11). The number in parentheses following the average values indicate the burning time in seconds over which the average was taken. Those following the maximum values indicate the time at which the maximum occurred. These values are from the nose-on aspect. Measurements were made at a range of 500 feet. No corrections have been made for atmospheric absorption. These measurements were made at the Naval Ordnance Test Station, China Lake, California, where the water content of the atmosphere is quite low.

TABLE IV.   Radiation Characteristics, Air-to-Air Rockets.

Type Rocket	1.8–2.8 μ		1.8–7 μ	
	average	max	average	max
FFAR MK 2 Mod 0	0.10 (1.6)	0.16 (0.7)	0.30 (1.9)	0.50 (1.0)
FFAR Experimental	1.2 (0.7)	2.2 (0.5)	1.0 (1.5)	2.0 (0.4)
FFAR 107 B	0.15 (1.6)	0.47 (0.3)	0.35 (1.8)	1.1 (0.9)
HVAR	1.0 (1.1)	1 7 (0.4)	3.2 (1.5)	6.0 (0.5)
Zuni	6.8 (1.2)	0.0 (0.3)	19.0 (1.6)	34.0 (0.9)
HPAG	0.05 (1.5)	0.30 (0.1)	0.16 (1.2)	0.28 (0.3)

TABLE 21-V.   Radiation Characteristics of Several Rockets
in the Ultraviolet

Type Rocket	Motor	Minimum	Maximum
FFAR	MK 1 Mod 3	0.1	0.13
FFAR	MK 2 Mod 0	3.0	8.0
FFAR	Exp X-12	12.0	15.0
HVAR	MK 10 Mod 6	18.0	27.0
HPAG	123 G	0.2	0.8

close to the nozzle, considerable energy at the centers of these bands is absorbed. Plume radiation is not particularly important in the 2.7 micron spectral region, it having been estimated that as much as 95 percent of the radiation in the lead sulfide region (1.8 to 2.8 microns) originates from the tail pipe. The chief significance of the plume radiation is that it has a much broader spatial distribution pattern than does the radiation from the tail pipe. There is considerably more energy from the plume in the 4.3 micron region of the spectrum than in the 2.7 micron region, although it is still much less than that of the tail pipe. The ratio of the output in the 4.3 micron band to that in the 2.7 micron band may vary by almost a factor of 10 and may be as much as 25 or 30. Since the 2.7 micron band radiation falls off more rapidly with temperature than does the radiation at 4.3 microns the ratio mentioned above increases with altitude, because of the lower throttle settings and correspondingly lower temperature usually involved. Total plume radiation drops rapidly with altitude. Measurements have shown that at 40,000 feet the plume radiation in the 2.7 micron region is down to about six percent of its ground level value, while the plume radiation in the 4.3 micron region is down to about 45 percent of its ground level value (Reference 9). The radiation in the 4.3 micron region at 40,000 feet would be about ten percent of the tail pipe radiation. At sea level the plume radiation of the J-57 engine in the region between 4 and 5 microns, for a throttle setting of 93 percent would be about 110 watts per steradian from tail aspect (Reference 3). At 40,000 feet this would drop to about 50 watts per steradian.

The infrared radiation from an aircraft is greatly increased by the use of an afterburner, the output of an engine sometimes increasing by a factor of 50 or more when its afterburner is turned on.

In the case of some supersonic aircraft the radiation from the aerodynamically heated skin will probably be much greater than that from the hot metal parts of the engines and the jet plumes combined. It has been estimated that a six-engine aircraft, flying at a speed of Mach 3 at an altitude of 75,000 feet may have a skin radiation of as much as 60 kilowatts per steradian for the 3.0 to 5.2 micron band in some directions if the skin material is in an oxidized condition (Reference 10).

Afterburner plume radiation is expected to be practically negligible compared to that of the skin. This is due to the fact that at these high speeds there is a great reduction in pressure as the hot gases leave the engine nozzle. This large change in gas pressure is accompanied by a corresponding drop in temperature. The resulting plume temperature behind the aircraft will probably be less than 500°C and the plume radiation in the 4.3 micron region will probably be less than 1 kilowatt per steradian, even when all six engines can be seen by the detector.

the same interval that the signal is presented. This simple statement is an example of the general proposition that for best results *the filter should be matched to the signal.* Under these circumstances there is output to the integrator only during the duration of the pulse, and noise occurring at all other times is eliminated. When supplemented with a frequency filter which limits the noise bandwidth to the smallest value consistent with the pulse duration, this time-domain filter gives results which approach the optimum predicted by information theory.

Time gates similar to the one just described are widely used in radar and communication systems. In ECM systems they are useful to sort out pulses received from various sources, whether the purpose be to secure information or to initiate countermeasures action.

The extreme simplicity of the range gate stems from the fact that the signal to be received consists of periodic rectangular pulses. In a more general situation the signal is modulated in both amplitude and phase, and it is necessary to employ considerably more complex filters in order to secure an appropriate match with the signal.

As an extreme example of a correlation filter, consider a c-w radar which transmits a signal consisting of bandlimited pure noise. Such a signal, which may be generated by passing pure "white" noise through a bandpass frequency-filter, consists of successive cycles, which while generally similar differ somewhat in both amplitude and period in a completely random manner. Because such a random sequence never repeats itself exactly, it is possible to secure range information by comparing the returned signal with a delayed sample of the transmitted signal. A method for performing this operation is shown in Figure 24-85.
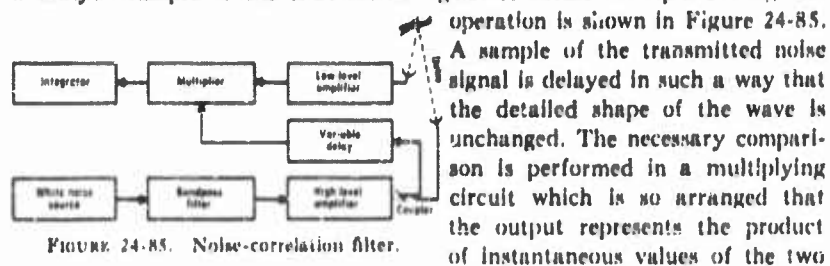
FIGURE 24-85. Noise-correlation filter.

A sample of the transmitted noise signal is delayed in such a way that the detailed shape of the wave is unchanged. The necessary comparison is performed in a multiplying circuit which is so arranged that the output represents the product of instantaneous values of the two inputs. This corresponds to the mathematical process of cross correlation. Under the specified conditions the output is extremely small unless the two inputs have the same waveform, in which case a substantial output is obtained.

It should be noted that the requirements on the delay unit for this application are much more severe than are those which apply to the simple range gate because it is necessary to preserve the wave shape rather than simply

produce a trigger pulse at an appropriate time.

The difficulty in producing a practical delay unit which meets the conditions specified is such that various artifices are employed to secure equivalent results. One such method is to recirculate the signal repeatedly through a relatively short delay line, compensating its losses by means of a broadband amplifier.

24.12.3 Comb Filters

Filters which freely transmit a large number of discrete frequencies which are harmonically related while suppressing all others are referred to as comb filters because the response curve looks like the teeth of a comb. While such filters can in principle be constructed of lumped elements using conventional design methods, it is rarely desirable to do so because of the large number of elements required and because the effects of dissipation are particularly harmful. A more rewarding approach is secured by noting that the desired spectrum represents the frequencies which are present in the most general form of wave having a period equal to the reciprocal of the desired frequency separation interval. Therefore, we are led to employ as our basic building block a passive unit having suitable values of bandwidth and time delay.
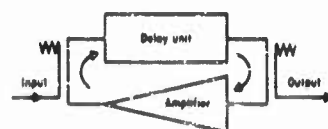
FIGURE 24-86.   Basic comb filter.

A comb filter of this sort is shown in Figure 24-86. The delay unit and amplifier have comparable and relatively large bandwidths, and the gain of the amplifier is only slightly less than the loss of the rest of the loop. Under these conditions the loop is strongly regenerative at a large number or uniformly spaced frequencies at which the net phase shift is an integral multiple of 360°. All such frequencies are strongly emphasized in the output. Therefore, the over-all transmission characteristic has the desired comb-like shape. The extent to which the ideal response may be approached depends upon the extent to which the loop gain at the many frequencies in question can be made to approximate the critical zero value.

The behavior of this type of comb filter and of all sorts of recirculation systems based on time delay is made clearer in the discussion of frequency memory in the following section.

The literature on filters is very extensive, and any brief bibliography necessarily represents somewhat arbitrary selection. A very complete theoretical treatment is given in Reference 63. Good handbook material is provided in References 64 and 65. A comparison of filtering methods is given in Reference 66.

24.13 Frequency Memory

In many ECM and ferret situations it would be convenient if a received signal pulse could be stored as a continuous wave of substantially the same frequency. This function is performed by a frequency memory unit, which is a special kind of regenerative circuit capable of sustaining oscillation at any one of a large number of discrete frequencies.

The arrangement of a typical frequency memory unit is shown in Figure 24-87. Delay, which is typically of the order of one microsecond, is provided by an appropriate coaxial cable. The traveling-wave tube, which has a bandwidth upwards of a thousand megacycles, provides a gain approximately 6 db in excess of the loss produced by the delay line and the directional coupler, which should have a relatively small coupling loss and very small reflection coefficients. The input signal divides in the coupler, part passing directly to the output, part flowing through the delay line to the amplifier. If the amplifier is energized in the presence of a relatively small input it will yield an additional component of output. Of greater present interest, it contributes to the delay line an additional signal which under favorable circumstances is in phase with and larger than the original signal. It is readily seen that the frequency of the signal tends to be preserved and that the level of the signal grows with successive recirculations around the loop. This step-wise growth process continues until a level is reached at which saturation sets in and the net gain around the loop is reduced to unity. Under favorable conditions the system now oscillates stably at this level and frequency until the power is turned off or the situation is changed by the injection of a relatively large input at some new frequency. The behavior is illustrated in Figure 24-88.

FIGURE 24-87. Frequency memory unit using traveling-wave tube and coaxial cable delay line.

FIGURE 24-88. Signal growth in idealized frequency memory unit with loop gain of 6 db.

It is readily seen that the possibility of oscillation exists for any frequency satisfying the Barkhausen phase relationship.

$$\phi = 2n\pi \ \text{(radians)} \tag{24-97}$$

If, as is often the case, the total loop delay is nearly independent of frequency and is equal to τ (seconds) the total phase shift is given by

$$\phi = \omega\tau = 2\pi/\tau \tag{24-98}$$

Eliminating ϕ between these equations we see that oscillation can occur at any frequency satisfying the equation

$$f = n/\tau \tag{24-99}$$

subject to the additional requirement that the (small signal) loop gain must equal or exceed unity. If, for example, the delay is one microsecond and the loop gain exceeds unity throughout the frequency band 3 to 4 kmc, then the unit is at least potentially capable of oscillation at any one of a thousand discrete frequencies separated by uniform increments of one megacycle.

The analysis (Reference 67) of frequency memory units, which is relatively complicated, shows that the idealized behavior just described is not realized in practice. Occasionally, the system oscillates simultaneously at two or more frequencies. More often, the small reflections produced by imperfections in the tube and cable disturb the form of the oscillation by producing growing phase modulation which eventually suppresses the carrier. When this occurs the initial frequency is lost and oscillation occurs at some frequency or frequencies favored by the system. However, the phase modulation process just described requires a substantial amount of time—therefore, even in unfavorable circumstances the initial frequency is preserved for some finite period referred to as the memory time.

As one might anticipate from simple considerations, the system works best at frequencies having the largest values of loop gain and more poorly at frequencies where the gain is lower. However, it turns out that the behavior is affected more by the curvature than the depth of the gain valley. This situation is illustrated by Figure 24-89 which shows the loop gain and memory time of a particular S-band frequency memory unit.

The behavior of a frequency memory system is greatly affected by the duration of the input pulse, which is often referred to as the instruction signal. It is readily seen that the optimum situation exists when the instruction pulse has a duration exactly equal to the loop delay τ and a frequency corresponding to one of the *natural modes* of the system as given by Eq (24-99). Finally, if the input level is such that the tube is immediately driven to its normal saturation level, there is no starting transient whatever. The

first cycle of the wave delivered from the output of the amplifier fits exactly onto the last cycle of the input pulse, and continuous oscillation is established.
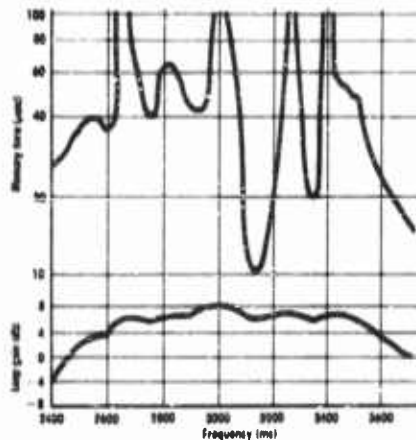
FIGURE 24-89. Loop gain and memory time for a typical frequency memory unit.

It should be noted that the arrangement shown has the advantage that the amplifier need not be fully energized until an interval τ after the beginning of the instruction pulse. This feature is important in practical systems because appreciable delays are involved in the gate circuits which perform this function. In other situations where the amplifier can be energized quickly enough, it is possible to reduce the required input signal (and the available output) by interchanging the location of the delay and the coupler with respect to the tube.

The understanding of frequency memory systems is facilitated by refering to the two mode oscilator illustrated in Figure 24-90 and studied by van
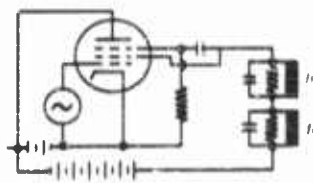
FIGURE 24-90. Bistable frequency memory using lumped constants.

der Pol (Reference 68). It is assumed that the tuned circuits are reasonably sharply tuned and are similar as to impedance level and natural frequency. If the two frequencies are incommensurate and the nonlinearity of the amplifier has a typical and appropriate form, it can be shown that the circuit can sustain oscillation at either of the two frequencies and that only one frequency is possible at one time. The frequency of oscillation can be established by an external generator connected in some convenient way as indicated. Two forms of instruction are possible. In one, the signal is applied during the interval immediately following the supply of power during which oscillation is building up. In the other a relatively strong signal is supplied for an interval sufficient to force oscillation at the desired frequency. Because the system can sustain oscillation indefinitely at either of two frequencies the choice of which is exterior to the system, it represents a form of memory unit operating in the frequency domain and capable of storing one bit of information.

By providing additional suitable tuned circuits to the arrangement of Figure 24-90 it is possible to extend the same kind of operation at three, four, or more separate frequencies. However, the impedance and frequency of the several tuned circuits must be controlled within narrower and narrower limits as the number is increased. Practical experience suggests that the system reaches its practical limit at about ten different frequencies.

24.14 Amplitude Limiters

In many systems, including automatic frequency control devices and receivers for frequency modulated signals, it is necessary to provide a circuit having an output which, within limits, is independent of the amplitude of the input. This function is performed by a suitable nonlinear circuit called an amplitude limiter.

One form of amplitude limiter which is applicable in a wide range of situations is the symmetrical clipper shown in Figure 24-91. The resistive load is approximately equal to the resistance of the source. The diodes are back biased so that they do not conduct until the instantaneous voltage exceeds
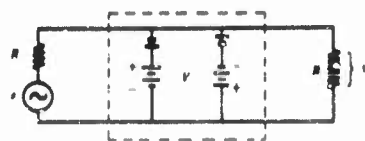
FIGURE 24-91. Symmetrical diode clipper.
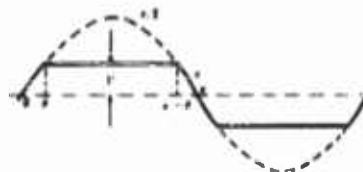
FIGURE 24-92. Waveforms in clipper.

some threshold level V. To facilitate analysis it is assumed that the diodes are ideal with zero forward resistance and infinite back resistance. Good results in substantial agreement with theory are obtained if in relation to the load resistance the forward resistance is very low and the backward resistance very high. Germanium, silicon, and thermionic diodes readily satisfy these conditions for reasonable values of the load impedance.

The action of the clipper is readily understood in terms of the waveforms shown in Figure 24-92. For small (instantaneous) values of e, the load voltage v is equal to $e/2$. For large values of e, the load voltage v is constant and equal to $\pm V$. It is seen that increase of the amplitude of e causes the output waveform to approach a square wave, of which the fundamental component has a (peak) amplitude equal to $\pi V/2$. A relatively simple analysis shows that the ratio of the fundamental component of the output voltage to the (sinusoidal) input voltage is expressed by the parametric equation

$$\gamma_2 = \frac{2(\theta + \sin \theta \cos \theta)}{\pi}$$

and has the form shown in Figure 24-93.

Amplitude Discrimination

Frequency modulation is widely used because it results in systems which are insensitive to important classes of noise and interference. This insen-
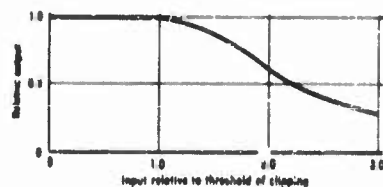
FIGURE 24-93. Radio of clipped to un-clipped output.
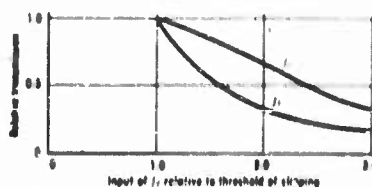
FIGURE 24-94. Amplitude discrimination in symmetrical clipper.

sitivity is directly connected with the *amplitude discrimination* provided by the limiter in the receiver. To examine this property we extend the analysis of the simple clipper of Figure 24-91 to include the effect of two signals. See Figure 24-94. Suppose that the source contains two unrelated frequencies f_1 and f_2 and that the voltage of f_2 is large compared to that of f_1. It is readily seen that the results previously presented still govern the behavior of f_2 but that f_1 is transmitted to the load only during the unclipped interval of the cycle. Thus the relative transmission of the f_1 component is given by the equation.

$$\gamma_1 = 2\theta/\pi$$

It is seen that the weaker signal f_1 is transmitted less freely than the strong signal at f_2. That is, the strong signal tends to suppress or *discriminate* against a co-existing weaker signal.

This desirable property can be compounded by using several limiters in cascade. However, it is necessary to separate successive clippers by appropriate tuned circuits of filters. Otherwise the clippers are effectively in parallel and no useful accumulation of the discriminating function results.

The cascading of limiters is greatly simplified by employing vacuum tubes to isolate successive stages. Successive tuned circuits compound their selectivity, and the vacuum tubes produce both gain and limiting. The ideas involved are well illustrated by the I-f amplifier, limiter arrangement for a frequency modulation receiver as illustrated in Figure 24-95. Each stage produces a relatively large gain of the order of 30 db for small signals. Thus
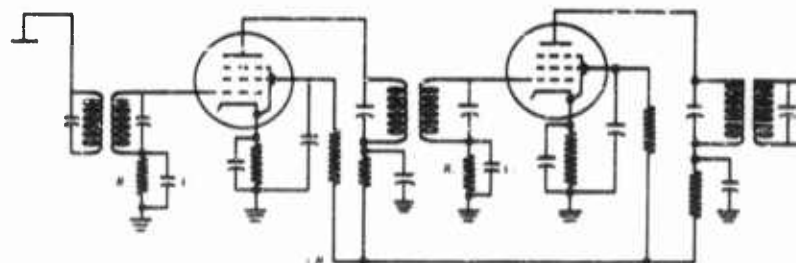
FIGURE 24-95.   Cascade pentode limiter as used in FM receivers.

a very small input produces an output sufficient to saturate the final tube, and a somewhat larger input saturates both of the last two tubes. However, the output changes very little with additional input because the signal driving the final tube is now limited and no further saturation is possible.

The limiting action results from rectification in the grid circuits which increases the bias and decreases the average plate current of the tubes as the signal input increases. If R_c is properly chosen with respect to the tube characteristics the plate supply voltage and the screen dropping resistance, the output voltage is substantially constant for a least a 10:1 ratio of input voltages. The associated capacitance C_c must be large enough to bypass R_c and should produce with it a time constant (Reference 69) compatible with the bandwidth of the signals to be received.

The bandwidth desired in the tuned interstage circuits depends upon the application. In automatic frequency control systems, it is not critical and need be merely sufficient to pass all the frequencies of interest. In receivers for frequency-modulated signals, the best signal-to-noise ratio is obtained if the bandwidth of these circuits is barely sufficient to pass the desired signals.

24.15 Superregeneration

Superregeneration is a term used to describe a form of signal reception which makes use of the exponential growth of an initially small signal in a circuit which has an effectively negative resistance, so that the logarithmic decrement is negative. Because limiting and saturation occur very quickly in a negative-resistance system, it is necessary to stop and restart the process at frequent intervals. The rate at which the starting process must be repeated is determined by the bandwidth of the signal to be received. Shannon (Reference 70) has shown that all the information conveyed in a frequency band extending from zero to frequency f is preserved if the wave is sampled periodically at frequency $2f$. For example, a speech channel having an upper frequency of 4 kc suffers no loss of fidelity if sampled at any rate greater than 8 kc.

The principal virtues of superregeneration stem from the fact that a very large gain is provided by a single tube and a high and controllable degree of selectivity is provided by a single tuned circuit. While it is possible to arrange matters so that a single tube produces the periodic sampling function as well as the negative resistance, it is rarely desirable to do so.
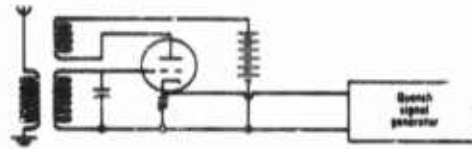
FIGURE 24-96. Superregenerative amplifier.

A superregenerative amplifier is indicated in Figure 24-96. A triode is coupled to an antenna and provided with feedback such that oscillation will occur when the bias is reduced below some threshold value. The quench signal generator produces a voltage which is periodic and has an appropriate waveform. Other waveforms are preferable, as indicated later, but the square wave is useful and serves to illustrate the basic idea.

During one-half cycle the tube is nearly or completely cut off, and any transient oscillations rapidly die away. Thus at the end of a very short interval the only voltage in the grid circuit is a measure of the antenna input signal, as affected by the selectivity of the resonant circuit During the next half cycle the bias is reduced and the plate current increases to a value such that regeneration is more than sufficient to overcome circuit losses. Now oscillations at the frequency of the tuned circuit build up exponentially with time from the level initially established by the antenna circuit.
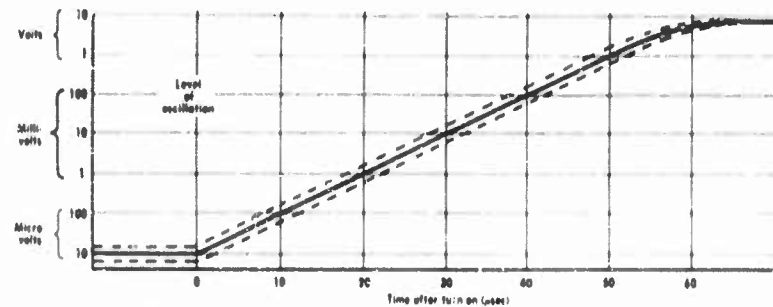
FIGURE 24-97. Exponential growth of signal development.

The basic action is shown in Figure 24-97 which plots the instantaneous level of oscillation against time, using a logarithmic ordinate scale. It is seen that the level increases by a very large amount (e.g. 100 db) in a relatively

short time (e.g., 50 μsec). The crucial fact is that until nonlinearity due to
saturation sets in, the voltage at the end of some definite time period is
(almost) proportional to the voltage which existed when the circuit was
activated. The approximation involved stems from the fact that the input
(antenna) signal not only affects the voltage at the moment of turn on, but
affects the growth rate for the first few microseconds thereafter.

At the end of the growth interval (e.g., 50 μsec), the oscillation level is
sampled by a simple detector or other appropriate means and the oscillation
is then damped out by turning off the tube. This action is shown in Figure
24-98. It is clearly necessary that the positive damping coefficient provided
during the turn-off interval is larger than the negative coefficient which exists
during the growth period. Otherwise, the oscillations built up during one
period persist or "hang over" into the next and the desired action is lost.
This undesirable circumstance is avoided by use of the special conductance
variation shown at the foot of Figure 24-98.
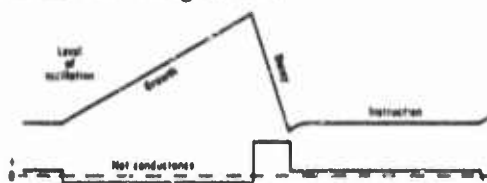
FIGURE 24-98. Complete cycle of operation.

The voltage level from which the next train of oscillations build up is
established during the period designated "instruction" just prior to the
moment of turn-on when the net conductance of the tuned circuit is made
negative. The instruction interval is identified as the sampling period in
terms of information theory.

Wheeler (Reference 71) has shown that the effective selectivity of a
superregenerative amplifier can be controlled between wide limits by shaping
the time-variation of conductance at the end of the instruction interval. If,
for example, the net conductance of the circuit is made nearly zero for a
relatively long time, the response is that of a single very high Q circuit.
Alternatively, a linear time variation of the conductance from a considerable
positive to a considerable negative value produces a selectivity curve of the
well-known gaussian ($y = ae^{-x^2}$) form. This is a remarkable result in that
the high skirt selectivity characteristic of the gaussian response is normally
attained only by cascading a large number of tuned circuits whereas it is
achieved here in a single circuit by shaping the pulse which controls the
gain of the tube and thereby the effective negative conductance of the tuned
system.

24.16 Locking in Oscillators

The terms *locking* and *synchronization* are used interchangeably to identify situations in which the frequency of an oscillator is controlled by the injection of a signal from some external source. In the simplest case, the frequency of the injected signal is nearly equal to the signal at which the oscillator would operate if undisturbed. When the difference between the input frequency and the free-running frequency becomes sufficiently small, it suddenly drops to zero and the oscillator acts as a highly regenerative amplifier of the input. The output consists of a single frequency which is identical with the input. That is, the output is *locked to* or *synchronized with* the input. The output is nearly constant as the input frequency is varied through the *range of synchronization* within which locking occurs, but the phase of the output with respect to the input varies through a range approaching $\pm 90°$.

Simple locking in a 1:1 frequency ratio occurs in all kinds of oscillators and at all frequencies where oscillation can be produced. In contrast, locking of a more complicated kind in which the input differs from the free oscillation frequency occurs only in nonlinear oscillators. This is not a serious practical limitation because most oscillators are quite nonlinear.

Ordinary oscillators consist of some sort of tuned circuit or resonator in combination with an electronic device which produces sufficient amplification or negative resistance to overcome the inherent losses. In certain microwave tubes, such as the backward-wave oscillator and the voltage-tuned magnetron, the tuning function is at least partially controlled by electron transit time. However, in all cases, the stable amplitude equilibrium which characterizes the state of sustained oscillation is associated with nonlinearity and saturation in the electronic device. The nature and extent of the nonlinearity of a particular oscillator governs the range of synchronization that a given signal will produce, but does not affect the nature of the phenomenon.

It is relatively easy to see that synchronization should occur when the synchronizing signal has a frequency which is half the free-running frequency of the oscillator. The injected signal, acting on the nonlinearity of the electron device, produces a second harmonic which may be thought of as locking the frequency on a 1:1 basis. Evidently, this line of reasoning can be extended to include any situation where the oscillator frequency is a simple harmonic of the locking frequency.

Locking also occurs when the synchronizing frequency is twice that of the oscillator. Perhaps the simplest way to view the situation is that the oscillator produces a substantial second-harmonic current which has the same frequency as the injected signal and thereby produces locking. Additional insight is gained if we note that the *difference* of the oscillator and input

frequency is equal to the oscillator frequency if and only if a two-to-one fre-
quency ratio exists. Analysis based on this concept predicts locking and
indicates the manner in which the phase varies as the input signal frequency
is varied within the synchronizing range.

Locking is observed whenever the oscillator and input frequencies are
related in the ratio of small whole numbers (such as 3:2). However, the
range of synchronization is relatively small, and the conditions for locking
are too critical to be of much practical value for any but the simplest fre-
quency ratios.

Locked oscillators have at least two important applications. In receivers
for frequency modulated signals, they combine the functions of amplification
and limiting. In frequency synthesis systems, they serve a filtering function
to produce a relatively large sinusoidal signal with a frequency identical
with some chosen component of a more complicated wave produced by
multiplication or modulation from some basic frequency source.

24.16.1 Range of Synchronization

The frequency interval over which an oscillator can be synchronized de-
pends upon the level of the synchronizing signal in comparison to the
magnitude of the self-generated oscillation, and upon the rate of change of
phase shift with respect to frequency in the resonator or other frequency
governing circuit. The essential facts of synchronization of a feedback
oscillator are shown in Figure 24-99. When the injected signal has a frequency
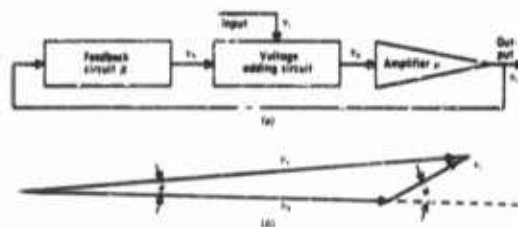
FIGURE 24-99. Synchronization of an oscillator. (a)
Block diagram. (b) Phasor diagram.

exactly equal to the natural frequency of the oscillator we see that v_i is in
phase with v_a and v_b, and that the angles θ and ϕ are both zero. At other
frequencies, the loop phase shift θ has a finite value and equilibrium requires
that ϕ have a larger value. At any given frequency, θ has a unique value
characteristic of the system, and it is clear that there is some minimum value
of v_i, corresponding to $\phi = 90°$, which will produce synchronization. From
this point of view it is clear that to secure synchronization over substantial

frequency intervals with relatively weak signals it is desirable to make the rate of change of the loop phase shift θ as small as possib' . This is achieved (Reference 72) by lowering the Q of the circuits involve' minimi ing stray capacitances to preserve adequate impedance levels, and by using special circuits which, (within the range of interest, have abnormally small variation of phase shift. The fact that the relative magnitudes of v_a and v_b must change somewhat as θ and ϕ vary, is identified with a small change of the level of oscillation and the degree of saturation in the amplitude limiter which occurs as the synchronizing frequency is varied.

REFERENCES

1. Terman, F. E., "Electronic and Radio Engineering," Sec. 9-2, 4th Ed., McGraw-Hill Book Company, New York, 1955.

2. Valley, G. E., Jr., and H. Wallman, "Vacuum Tube Amplifiers," Chap. 2; MIT Rad. Lab. Series, McGraw-Hill Book Company, New York, 1948.

2a. Pettit, J. M., and M. M. McWhorter, Electronic Amplifier Circuits, McGraw-Hill, New York, 1961.

3. Palmer, R. C. and L. Mautner, A New Figure of Merit for the Transient Response of Video Amplifiers, Proc. IRE, Vol. 37, pp. 1073-1077, September 1949.

4. Easton, A., Stagger Peaker Video Amplifier, Electronics, Vol. 22, p 118, February 1949.

5. Mulligan, J. H., Jr. and L. Mautner, Steady-state and Transient Response of Feedback Video Amplifiers, Proc. IRE, Vol. 36, pp. 545-610, May 1948.

6. Elmore, W. C., The Transient Response of Damped Linear Networks with Particular Regard to Wideband Amplifiers, J. Appl. Phys., Vol. 19, pp. 55-63, January 1948.

7. Mathers, W. C., Ph.D. dissertation, Stanford University, 1951.

8. Wheeler, H., Wideband Amplifiers for Television, Proc. IRE, Vol. 27, p. 429, 1939.

9. Bedford, A. V. and G. L. Fredendall, Transient Response of Multistage Video-frequency Amplifiers, Proc. IRE, Vol. 27, p. 277, April 1939.

10. Kallman, H. E., R. E. Spencer, and C. P. Singer, Transient Response, Proc. IRE, Vol. 33, pp. 169-195, March; p. 482, July; 1945.

11. Tucker, D. G., Bandwidth and Speed of Build-up as Performance Criteria for Pulse and Television Amplifiers, Proc. IRE, Vol. 04, pp. 218-226, May 1947.

12. Elmore, W. C., Electronics for the Nuclear Physicist, Nucleonics, Vol. 2, pp. 4-17, February; pp. 16-36, March; pp. 43-55, April; pp. 50-58, May; 1948.

13. Elmore, W. C. and M. Sands, "Electronics: Experimental Techniques," Chap. 3, National Nuclear Energy Series, McGraw-Hill Book Company, New York, 1949.

14. Richards, P. I., Cathode-Follower Fallacies, Rev. Sci. Instr., Vol. 21, p. 1026, December 1950.

15. Bode, H. W., "Network Analysis and Feedback Amplifier Design," D. Van Nostrand Company, New York, 1945.

16. Guilleman, E. A., Mathematics of Circuit Analysis, pp. 330-349, John Wiley and Sons, New York, 1950.

17. Terman, F. E., "Radio Engineering," 3rd Ed., McGraw-Hill Book Company, New York, 1947.

18. Nuttall, T. C., Some Aspects of Television Circuit Technique: Phase Correction and Gamma Correction, *Bull. Assn. Suisse Electr.*, Vol. 40, pp. 615-622, August 1949.

19. Gourlet, G. G., V. F. Amplifier Couplings, *Wirel. Engr.*, Vol. 27, pp. 257-265, October- November 1950.

20. Brain, A. E., The Compensation for Phase Errors in Wide-Band Video Amplifiers, *Proc. IER, (London)*, Part III, Vol. 97, pp. 243-251, July 1950.

21. Percival, W. C., "Thermionic Valve Circuits," British Patent 460562, July 24, 1935-January 25, 1937.

22. Ginzton, E. L., W. R. Hewlett, J. H. Jasberg, and J. D. Noe, Distributed Amplification, *Proc. IRE*, Vol. 36, pp. 956-969, August 1948.

23. Payne, D. V., Distributed Amplifier Theory, *Proc. IRE*, Vol. 41, pp. 759-762, June 1953.

24. Scarr, R. W. A., *Proc. IRE*, Vol. 42, pp. 596-598, March 1954.

25. Copson, A. P., A Distributed Power Amplifier, *Elec. Eng.*, Vol. 69, pp. 893-898, October 1950.

26. Kallman, H. E., Equalized Delay Lines, *Proc. IRE*, Vol. 34, p. 646, September 1946.

27. "Waveforms," MIT Rad. Lab. Series, pp. 730-750, McGraw-Hill Book Company, New York, 1949.

28. "Components Handbook," MIT Rad. Lab. Series, pp. 191-217, McGraw-Hill Book Company, 49.

29. Trevor, B., Jr, Artificial Delay-Line Design, *Electronics*, Vol. 18, p. 135, June 1945.

30. Murphy, B., Distributed Amplifiers, *Wirel. Engr.*, Vol. 30, pp. 39-47, February 1953.

31. Rudenberg, H. G. and F. Kennedy, 300 Mc Travelling-Wave Chain Amplifier, *Electronics*, Vol. 22, pp. 106-109, December 1949.

32. Horton, W. H., J. E. Jasberg, and J. D. Noe, Distributed Amplifiers, Practical Considerations and Experimental Results, *Proc. IRE*, Vol. 38, pp. 748-753, July 1950.

33. Linvill, J. C., Amplifiers with Prescribed Frequency Characteristics and Arbitrary Bandwidth, Tech. Report No. 163, Research Laboratory of Electronics, MIT, July 7, 1950.

34. Wheeler, H. A., Maximum Speed of Amplification in a Wideband Amplifier, Wheeler Monograph No. 11, Wheeler Labs., Inc., Great Neck, New York, July 1949.

35. Butterworth, S, On the Theory of Filter Amplifier, *Exp. Wirel. and Wirel. Engr.*, Vol. 7, pp. 536-541, October 1930.

36. Dishal, M., Design of Dissipative Band-Pass Filters Producing Exact Amplitude-Frequency Characteristics, *Proc. IRE*, Vol. 37, pp. 1050-1069, September 1949.

37. Dishal, M., Alignment and Adjustment of Synchronously Tuned Multiple-Resonant-Circuit Filters, *Proc. IRE*, Vol. 39, pp. 1448-1455, November 1951.

38. Whyte, J. R., Choosing Pentodes for Broad-Band Amplifiers, *Electronics*, Vol. 25, p. 150, also Fig. 2, April 1952.

39. Landon, V. D., "Cascade Amplifiers with Maximal Flatness," *RCA Rev.*, Vol. 5, pp. 347-362 and 481-497, January and April, 1941.

40. Lynch, W. A., The Role Played by Derivative Adjustment in Broadband Amplifier Design, *Proc. Symp. on Mod. Netw. Synthesis*, pp. 193-201, Polytech. Inst., of Brooklyn, New York, 1953.

41. Thomson, W. E., Networks with Maximally Flat Delay, *Wirel. Engr.*, Vol. 29, pp. 256-263, October; p. 309, November; 1952.

42. Muller, F. A., High-Frequency Compensation of RC Amplifiers, *Proc. IRE*, Vol. 42, pp. 1271-1275, August 1954.

43. Bradley, W. E., Design of a Simple Band-Pass Amplifier with Approximate Ideal Frequency Characteristics, *IRE Trans.*, PGCT-2, pp. 30-38, December 1953.

44. Schienemann, R., "Trägerfrequenzverstärker grosser Bandbreite mit gegeneinander verstimmten Einzelkreisen," *Telegraphen Fernsprech Technik*, pp. 1-7, 1939.

45. Wallman, H, Stagger Tuned I-F Amplifiers, MIT Rad. Lab. Report No. 524, February 1944.

46. Baum, R. F., Design of Broad-Band I-F Amplifiers, *J. Appl. Phys.*, Vol. 17, pp. 519-529 and 921-930, 1946.

47. Trautman, D. L., Jr., Maximally Flat Amplifiers of Arbitrary Bandwidth and Coupling, Tech. Raport No. 41, ERL, Stanford, February 1, 1952.

48. Eddy, J. S,. Stagger Tuned Amplifiers with Double-Tuned Interstages, Tech. Report No. 29, ERL, Stanford, January 1951.

49. Trautman, D. L. and J. A. Aseltine, Equal-Ripple Bandpass Amplifiers, Report 51-9, Dept. of Engin., University of California, Los Angeles, August 1951.

50. McWhorter, M. M., The Design, Physical Realization and Transient Response of Double-Tuned Amplifiers of Arbitrary Bandwidth, Tech. Report No. 58, ERL, Stanford, February 1953.

51. Pettit, J. M., Amplifier Circuit Theory, Stanford Bookstore, 1957.

52. Wittenberg, R. C., Broad-banding by Stagger Tuning, *Electronics*, Vol. 25, pp. 118-121, February 1952.

53. Jenkins, E. R., Stagger Gain Calculator, *Tele-Tech*, Vol. 9, p. 29, April 1950.

54. Wightman, B. A., A Graphical Method for Determining the Number and Order (N) of N-uples in Stagger Tuned Amplifier Design, Report ERA-212, National Research Council of Canada, December 1951.

55. McWhorter, M. M. and J. M. Pettit, The Design of Stagger-Tuned Double-Tuned Amplifiers for Arbitrary Large Bandwidth, *Proc. IRE*, Vol. 43, pp. 923-931, August 1955.

56. Edson, W. A., The Single-Layer Solenoid as an RF Transformer, *Proc. IRE*, Vol. 43, pp. 932-936, August 1955.

57. Kinaman, E. W. and M. Magid, "Very Low-Noise Traveling Amplifier," *Proc. IRE*, Vol. 46, No. 5, pp. 861-867, May 1958.

58. Fank, F. B. and F. M. Schumacher, "Development and Operation of Broadband Low-Noise Traveling-Wave Tubes for X and C Bands," IRE Wescon Convention Record, Part 3, Electron Devices, pp. 150-155, 1957.

59. *Electrical Design News*, pp. 60-61, January 1960.

60. Brand, F. A., W. G. Matthei, and T. Saad, The Reactron—A Low-Noise, Semiconductor Diode, Microwave Amplifier, *Proc. IRE*, Vol. 47, No. 1, January 1959.

61. Bloom, S., and K. K. N. Chang, Theory of Parametric Amplification Using Nonlinear Reactances, *RCA Rev*, Vol. 18, pp. 578-593, December 1957.

62. Shea, T. E., "Transmission Networks and Wave Filters," D. van Nostrand Company, 1929.

63. Guillemin, E. A., "Communication Networks," Vol. 2, John Wiley & Sons, Inc., New York, 1935.

64. "Reference Data for Radio Engineers," 4th ed., International Telephone and Telegraph Co., New York, 1956, pp. 164-246.

65. Pender, H., and K. McIlwain, "Electrical Engineers Handbook," 4th ed., John Wiley & Sons, Inc., New York, Sec. 6, pp. 33-62.

66. George, S. F., "Time Domain Correlation Detectors vs. Conventional Frequency Domain Detectors," NRL Report 4332, Naval Research Laboratory, Washington, D.C., 1954.

67. Edson, W. A., Frequency Memory in Multi-Mode Oscillators, *Trans. IRE*, CT 2, March 1955.

68. van der Pol, B., The Non-Linear Theory of Electric Oscillations, *Proc. IRE*, Vol. 23, pp. 1051-1086, 1934.

69. Baghdady, E. J., FM Demodulator Time-Constant Requirements for Interference Rejection, *Proc. IRE*, Vol. 46, pp. 432-440, February 1958.

70. Shannon, C. E., A Mathematical Theory of Communication, *Bell System Tech. J.*, p. 379, July; p. 623, October; 1948.

71. Wheeler, H. A., "A Simple Theory and Design Formula for Super-regenerative Receivers," Wheeler Monographs, Vol. 1, Wheeler Laboratories, 1953.

72. Carnahan, C. W. and H. P. Kalmus, Synchronized Oscillators as F-M Receiver Limiters, *Electronics*, Vol. 17, pp. 108-111, August 1944.

25

Mechanically Tuned High-Power Oscillators and Amplifiers

R. B. NELSON, P. W. CRAPUCHETTES

Progress in the design of components for electronic countermeasures (ECM) has historically moved ahead in bursts followed by long pauses. In no other ECM component has this been more true than in the design of tubes for use in mechanically tuned oscillators and amplifiers. ECM had its beginnings with the design of systems around commercially available triodes, originally designed for fixed-frequency studio-transmitter link service. Newer systems were designed concurrently with new tube types—planar triodes, resnatrons and split-anode magnetrons. These systems were all operational in World War II. Toward the close of the war the ECM systems using multi-cavity magnetrons were studied extensively and scientifically. Particular attention was paid to considerations determining jamming effectiveness. These studies continued in the period immediately following World War II, culminating in the several magnetron systems which currently provide the ECM capability of the Armed Forces. Because of its several desirable characteristics, the floating drift tube klystron was also extensively studied in this period.

The mechanically tunable oscillator has tended to fall into disrepute in an era of sophisticated radars having new antijam features, such as multi-plex, frequency diversity, and shaped pulses. The multiple task requirements tended to make brute force too large and too heavy. Lately, however, a new capability has come into being in the use of mechanically tuned oscillators. The use of increased bandwidth barrage self-noise or video-enhanced noise

In mechanically tuned oscillators improves their usefulness against the more sophisticated radars.

25.1 Summary of the Relative Operating Parameters of the Various Devices

The relative merits and demerits of the tube types are described in the following tables. Table 25-1 describes the present status and essential charac-

TABLE 25-1. Application Comparisons of the Various Tubes

Feature	Triode	Tetrode	Reinatron	Spill-Anode Magnetron	Multicavity Magnetron	Floating Drift Tube Klystron
Circuit	tptk[a]	tptk[a]	tptk[a]	lecher line	internal	internal
Unit size	bulky	bulky	bulky	bulky	compact	compact
Accessories	b, c	b, c	b, c	d, e	c, e	e
Tuning knobs	2[f]	2[f]	2[f]	1[g]	1	1[h]
Tuning ease	fair	fair	hard	easy	excellent	good
Tuning rate	slow	slow	0	slow	fast	medium
Reliability	good	good	fair	good	excellent	lab.
Life (hr)	50-250	50-250	8-100	50-250	300-1000	5000
Availability	prod.	prod.	lab.	prod.	prod.	lab.
Amplification Capability	yes	yes	yes	no	no	no
Usage	Obs.	Obs.	Obs.	Obs.	99%	lab.

[a] Ground grid circuit, tuned plate and cathode normally used.

[b] Filament schedule required to compensate for back-heating.

[c] Circuit not self-stable, use cathode self-bias.

[d] Filament regulator required for long cathode life.

[e] Magnetic field required, may be permanent magnet.

[f] Two-cavity circuits used with feedback and output circuits controlled by adjustable rolling cams to accept tube variations.

[g] Load vernier needs adjustment after large frequency change.

[h] Voltage must be tracked by servo while tuning.

teristics of ECM systems using the various types of mechanically tuned devices. It indicates that, excepting the magnetron, all the various mechanically tuned devices which have been in service are now obsolete. Obsolescence has been caused by the introduction of tunable radars which impose a rapid tunability requirement upon the ECM system. Even the magnetron will be obsolesced soon except for spot jamming of particular situations. The

Table 25-2. Jamming Characteristics of the Devices

	Max. Bandwidth [a] (% f_0)	Max. Video [b] (% f_0)	Cavity Q	Z Mod [c] R	Z Mod [c] C	Composition of Jamming Signals
Negative grid tubes	1	0.5	30-100	1 to 10	15 to 100	Less than 1/2 radiated power is effective since carrier does not jam. Tetrodes are least effective.
Multicavity magnetrons	0.2 to 1	0.1 to 0	100 to 300	0.100 to 1.0	15 to 30	Correlated FM and AM result of pushing inherent to the tube except below P band.
Barretron[d]	5.0	5.0	10 to 30	about 0.3	15 to 30	Wideband self-noise with little correlation between AM and FM. May be video modulated for more bandwidth.
Floating drift tube klystrons	1.0	0.5	100	AM FM		Composition can be selected at will from full AM to full FM to any correlated or uncorrelated combination.

[a] Spectrum bandwidth is normally twice video unless pushing is present, in which case the spectrum widens. Task circuit Q limits the spectrum to the value shown, regardless of modulation frequency.

[b] Video components are generally equal to the modulation frequency which can be equal to or less than the given percent of carrier frequency, f_0.

[c] Z mod. is the modulated electrode impedance in kilohms and microsiemens.

[d] Trade mark registered.

25-3

TABLE 25-3.   Equipment Usage by Various Tube Types

Triodes	AN/APT-5, "carpet-bag," AN/APT-9
Tetrodes	Project Tuba, APT-1 & 3
Split-anode magnetrons	TDY, AN/APT-2
Multicavity magnetrons	AN/XMBT-1 & 2, AN/MLQ-2 & 7, AN/ALT-2, 5, 8, AN/ALQ-23, AN/APT-16
Floating drift tube klystrons	Experimental
Barratrons*	Under consideration are AN/ALT-2, 6, AN/ALQ-23

Barratron * is a form of mechanically tuned self-noise generator which may offer some improvement over the magnetron.

Table 25-2 sets forth the relative capability of the various tube types in their conventional circuits as generators of microwave jamming energy. Table 25-3 lists some of the older equipments for which the various types of tubes were used.

25.2 Multicavity Magnetrons in Mechanically Tuned Oscillators

Single-dial control of frequency at rapid mechanical rates has made the multicavity magnetron the most popular of the mechanical ECM devices. Figures 25-1 and 25-2 indicate typical production devices. The addition of a moving magnetic member to the moving tuning structure has made possible wide tuning range (1.7:1 at P-band to 1.2:1 at X-band) while operating at essentially constant voltage. A tetrode modulator in a series-shunt arrangement with the magnetron cathode not only provides modulation but serves to hold current constant as well. Auxiliary circuits are available and reliable. Output coupling is fixed by the designer so that good performance is obtained over the desired band without adjustment in the field. High-quality noise is available from these high-efficiency devices because of their unique non-linear nature. Following a brief summary of magnetron design practice, there will be several paragraphs relating to the normal and abnormal characteristics of magnetrons plus further details of performance.

25.2.1   Multicavity Magnetron Design Summary

A magnetron is a diode with an axial magnetic field parallel to the axis of the cathode and anode space. In multicavity magnetrons, the anode cylinder is divided into uniform segments which are connected to resonant circuits that provide rf impedance, store energy, and regulate frequency. A physical picture of the electron interaction in magnetrons will materially assist the user in assuring their proper operation. The extensive literature

*Trade mark registered.

available mainly deals with pulse magnetrons, which differ in only small detail from the ECM magnetrons, largely in cathode design and output coupling circuits. References 1, 2, and 3 provide a very detailed picture of the electronics as well as excellent introductory material. Cathodes of ECM magnetrons are invariably fabricated of refractory metals, in order to sustain without damage the relatively large number of electrons which bombarded it from the electron stream. The cathodes must provide primary emission, whereas pulse-type magnetrons operate satisfactorily with secondary emission. The structures must be rigid and suited for operation under severe environmental conditions.

FIG. 25-1. QK-496 200-watt CW magnetron, hydraulically tunable from 2350 to 3600 Mc. Position controlled by servo. Maximum tuning rate: 20 cps over any 200-Mc portion of band. Body is liquid cooled. Tuner is conduction cooled (see Figure 25-4). *Courtesy Raytheon Manufacturing Co.*
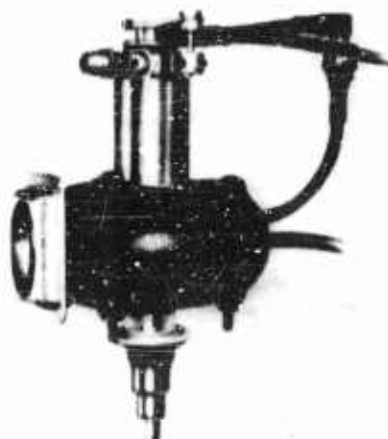
FIG. 25-2. L-3110A magnetron, 200-watt CW, tunable 2350-3600 Mc, showing separate tuner and body cooling circuits. Hydraulic tuner is servo actuated at 20 cps maximum rate over 400-Mc band. *Courtesy Litton Industries.*

The output circuit of an ECM magnetron is generally more heavily coupled to the resonant circuit than in pulse tubes so that an adequate noise bandwidth can be obtained (efficiency is also increased). The tube incorporates such combinations of transformers and discontinuities as are required to make coupling adjustment during operation unnecessary. The magnetic field required to provide the necessary synchronous velocity is readily provided by magnets which are permanently attached to the tube. Removal of the

magnets will destroy operation unless the magnets are remagnetized after reassembly. To reduce size and weight and stray magnetic-field leakage, some magnetrons are packaged with bowl magnets.

25.2.1.1   Interaction Space Design. The basis for much of the interaction space theory for magnetrons was laid by A. W. Hull (Reference 4), who investigated the behavior of electrons in a cylindrical diode in the presence of an axial magnetic field. An electron leaving the cathode is acted upon by a radial force proportional to plate voltage and by a force acting at right angles to its velocity whose magnitude is proportional to the velocity and the d-c axial magnetic field strength, B. This combination of forces causes the electrons to move across the interaction space in quasicycloidal orbits. When the orbit just touches the anode, any further increase in magnetic field will turn the electrons away from the anode and a cutoff condition exists, as given by equation

$$V_{eo} = \frac{eB^2 r_a^2}{8m}\left[\ 1 - \left(\frac{r_c}{r_a}\right)^2\ \right]^2 \tag{25-1}$$

where r_a and r_c are anode and cathode radii, respectively, and V_{eo} is the voltage from anode to cathode at the d-c cutoff condition.

Several alterations of structure were found to yield oscillation in the near-cutoff region. Further study showed that, if the cylindrical anode were divided equally into an even number (N) of parts (see Figure 25-3), and

Fig. 25-3. Plan view of a magnetron showing vanes and straps. *Courtesy Litton Industries.*

suitable impedances were introduced in consecutive array, then large signal oscillation would be observed. Current is drawn to the anodes under the combined influence of the applied voltage and the rf voltage, even though the applied voltage is much less than the d-c cutoff value from Equation 25-1. The standing wave of voltage produced on the vane tips (which are of alternating polarity in normal operation) can be separated into two oppositely traveling waves which must then be synchronous with the electrons whose velocity is determined by the geometry and the impressed

fields. O. Bunneman, working in a group under Hartree (Reference 5), developed the criteria determining the voltage which must be applied to a given geometry if oscillations are to start, which is given by

$$V_{op} = \frac{2\pi}{N\lambda} \, (r_a{}^2 - r_c{}^2) \left(B - \frac{\pi m c r_a{}^2}{e N \lambda} \right) \tag{25-2}$$

in which λ is wavelength and V_{op} is the voltage at which oscillations start provided that B is large enough so that no current will be drawn to the anode in the absence of oscillations. A further study of these two relations by D. A. Wilbur (Reference 6) has permitted deduction of the electronic efficiency by

$$n_e = 1 - \frac{V_{op}}{V_{eo}} \tag{25-3}$$

The results agree well for low values of efficiency. When V_{eo}/V_{op} is greater than 1.5, it has been found that a factor of 2 in the denominator yields better agreement with the data.

Experience has dictated that the interaction space design must concern itself with two more parameters. The anode-to-cathode distance has been found to influence starting time, the dynamic impedance of the tube, and electron back-bombardment of the cathode. A reduction in spacing reduces all but the latter factor because of the greater projection of the anode fields toward the cathode. If the spacing is made too small, the magnetic-field requirement becomes uneconomic and back heating becomes destructive.

The ratio of the vane thickness to gap between vanes determines the composition of the fields in the interaction space since the magnitude of the various space harmonics will be determined by this ratio. At a value of 2:1 the fields are largely first harmonic and thus provide the greatest field for interaction per volt between vanes. The improvement in starting time which results is accompanied by a more uniform change in frequency due to increase in space charge when the current is increased. This is called frequency pushing. For further discussions of pushing, see Sections 25.7 and 25.9.

The three relations useful in magnetron design given above incorporate the variables V_{op}, V_{eo}, N, λ, n_e, r_a, r_c, and B, which means that the designer must draw on experience and, by successive approximation and selection of values for five of the variables and computation of the balance, determine a reasonable design.

25.2.1.2   Tuning. In the earlier paragraph introducing the synchronism criteria, the resonant frequency was described as a function of the induc-

tance and capacitance of the anode. Any variation of these parameters will cause resonant frequency to change. Typical tuning structures which modify capacitance and inductance are shown in Figures 25-4 and 25-5. Tuning

FIG. 25-4. Subassemblies of QK-301A. Note particularly the cooling of the tuners by flexible copper straps. Note the use of insulated end shields to increase stability. *Courtesy Raytheon Manufacturing Co.*

FIG. 25-5. Tuning structure for L-3110 magnetron. Top part is moving pole piece, next is capacitive tuner ring, then inductive tuner. Tungsten rods wiping on copper are used as line up bearings in the vacuum. *Courtesy Litton Industries.*

capacitively is accomplished by insertion of the smaller metallic member into the region of high E field near the tips of the vanes, effectively increasing capacitance by so doing. Inductive tuning is accomplished by inserting the larger metallic member into a region of high H, where the tuner currents oppose the incident field and force the field lines out of the tuned region, thus reducing the inductance. Combinations of E and H tuners (sometimes called L and C tuners) with moving pole pieces can be used to greatly increase tuning range. The moving pole piece is positioned and sh ed so that the proper variation in B as required by Equation 25-2 can b. approximated. Exact compensation is difficult since the tuning curve is essentially linear and the curve of B versus pole-piece position is hyperbolic. Over the usual tuning ranges of magnetrons, this compensation is adequate and the resultant loss in the current regulating circuit is minimized.

25.2.1.3 Family Capabilities.   The wide selection of operating para-

meters left to the designer makes possible the design of electrically and mechanically interchangeable tubes over a very wide band of frequencies in which only the rf output connection need be selected according to frequency. Illustrating these output configurations, Figure 25-6 shows part of a family

Fig. 25-6. Portions of CW magnetron families indicating output types. Fluid-cooled family, 200-watt CW; air-cooled family 100-watt CW. *Courtesy Litton Industries.*

of ten types of magnetrons which provide jamming from 350 to 10,500 Mc. The family operates from one power supply and provides good conversion efficiency throughout the range, along with small size and weight.

25.2.1.4   Secondary Electronic Effects in the Interaction Space. References 1 and 3 include illustrations which show some interesting details of the electron transit in the interaction space. Electrons leave the cathode uniformly over its surface. The action of the combined d-c and rf field tends to sort them into spokelike streams moving outward toward the anode while traveling synchronously around it. Some of the electrons are rejected in this forming process and are returned to the cathode as the result of their having entered the space-charge cloud in an unfavorable phase. In this circumstance the electron absorbs energy from the rf field and strikes the cathode with a velocity determined by that energy, overheating the cathode. When the cathode is close to the anode, the stream is more completely defined, encompassing a smaller phase angle and the number of rejected electrons and their energy increases.

The circulating cloud of electrons interacts with the total field in the interaction space. The d-c field is uniformly present, but there are two components of traveling rf wave in the interaction space, one having a radial direction vector and the other a tangential vector. These rf fields are spatially separated by one-half the pitch between vanes, since the one is derived from the vane-to-cathode standing waves which are maximum under a vane, and the other is derived from the vane-to-vane voltage which is a maximum in the space between vanes. Since the resonance criteria must be adhered to at all current levels, the input voltage of the magnetron remains relatively constant at all values of current and the rf voltage increases only as the square root of current input. Thus as operating level is changed the ratio of the number of electrons to the magnitude of the rf fields is changed; and oscillations can continue only if the phase of the spoke of the bunch relative to the rf field changes. With a change in phase of the bunch relative to the spoke, there is a change in coupling to each of the traveling fields and thus reactive components of currents must be induced upon each wave component. Therefore the frequency of the magnetron is a function of current; and this phenomena is called *pushing*.

When the phase of the spoke is such as to be coupling a maximum of energy from the electron stream, any attempts to further increase current cannot be supported by the net increase in rf field and the bunch collapses. This condition known as *mode shift* can be separated from the other causes for electron instability by suitable instrumentation since the dynamic impedance of the tube remains continuous up to mode shift. In Figure 25-7a an oscilloscopic plate diagram shows this mode boundary.

Mode shift can be induced by inadequate cathode emission. Because the supply of electrons is not equal to the demand, the plate voltage increases, curving upward, causing the velocity of the electrons to increase out of synchronism, and the bunch collapses, as illustrated in Figure 25-7b.

Some coupling between resonators generally occurs in multicavity magnetrons. The response of the coupled system as a function of frequency is split up into $N/2$ modes or resonances. To keep the separation in frequency of these modes great enough the coupling is generally made very tight through the use of the concentric rings called straps, which are connected to alternate vanes, as shown in Figure 25-3. When, however, the resonance criteria of Equation 25-2 can be satisfied equally well by the $N/2$ (π) mode and by the $N/2 - 1$ mode, the device will not simultaneously operate in both modes but will, in accordance with the Principle of Least Work, select that mode which produces greatest entropy, always $N/2 - 1$, since it is not

symmetrical. When the tube shifts from π mode because of competition, the operating conditions for the new mode (higher V_{op}, lower current) can be used to determine cause. Diagnosis of the cause for mode shift can usually be made if adequate test equipment is available.

(A) (B)
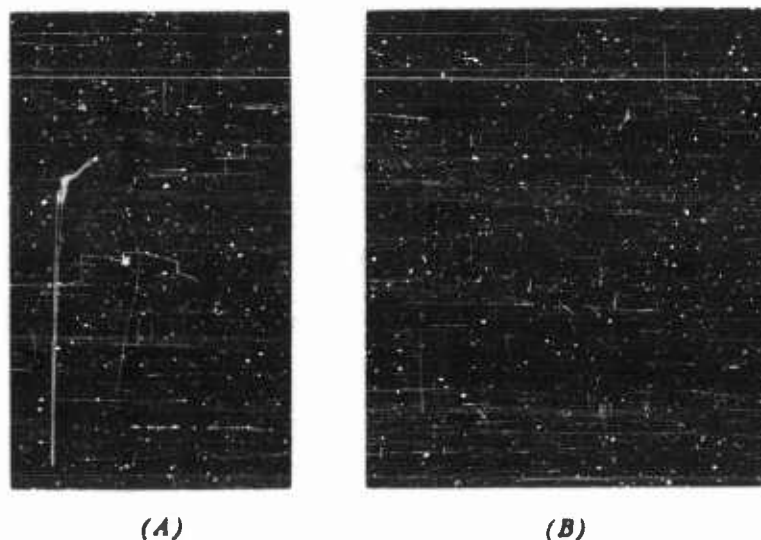
Fig. 25-7.   Plate diagrams (a) Mode shift; tube is being modulated at 60 cps to illustrate the phenomenon e_b is vertical; i_b is horizontal. (b) Mode collapse due to lack of emission. Note upward curl at end of π mode trace.

25.2.2   Operating Characteristics of Multicavity Magnetrons

The spectrum produced by an ECM magnetron is a function of the design of the magnetron and the design of the modulation equipment, the output line and the antenna. The plate diagram of a magnetron is the equivalent of that of a biased-off diode, in which the bias represents the start of oscillation voltage. Since a change in current produces a change in frequency, it is necessary that the magnetron current be stabilized, especially since any small change in supply voltage tends to produce a large change in current.

Tetrode current regulators have been found adequate for this function and can also serve as the video modulator. The upper modulation frequency is limited by anode Q and the capability of the modulation network. The rf bandwidth is at least as great as the video bandwidth and may be increased by pushing when modulated or by pulling when the antenna introduces a reflection of varying phase.

25.2.2.1 Pushing. In the section on secondary effects in the interaction space the phase of the space-charge wheel of spokes was shown to vary with anode current. Pushing, the change in operating frequency produced by a change in operating current in a specific way, usually is measured from 25 to 100 percent of rated peak current. The change in the number of electrons and the phase of the bunch produces an increase in frequency as current increases. For particular cases of loading the sense sometimes reverses; the turn-around or point of zero slope of the pushing curve increases with an increase in loading.

25.2.2.2 Pulling. The resonant frequency of an anode is determined by considering the effects of all the impedances in the anode circuit. When the load is reactive, it contributes to the total circuit reactance and thus causes frequency to be dependent on loading, both in magnitude and phase. *Pulling* is that change in operating frequency which results when a 1.5:1 mismatch is varied through all its phases. One should observe that *pushing* is the frequency change caused from effects originating within the magnetron and *pulling* is caused from effects originating without the magnetron.

25.2.2.3 Long-Lines Effects. When the magnetron is coupled to a load having some mismatch with the line impedance, the length of the line and the magnitude of the mismatch working together can cause holes in the spectrum or nonoperating regions. The effect of the line which is n wavelengths long may be so phased as to result in a negative pulling from the matched frequency. If the line is long enough so that $n - \frac{1}{2}$ wavelengths can be satisfied by an equal positive pulling shift, then the tube will operate at either of these points but never at intermediate points. This phenomena has been extensively studied and reported on by J. F. Hull, G. Novick, and R. Cordray (Reference 7). The addition of a load isolator near the generator will minimize or eliminate this problem. The paper by Hull *et al.* can be used to determine the one-way insertion loss required for hole-free operation with a line of any length and any attenuation.

25.2.2.4 Thermal Drift. Since the magnetron tuners are in general

supported and positioned by a long series of metallic members, the ultimate position of the tuner under varying ambient conditions is variable. Proper cooling of the tuner parts and support will minimize the shift in frequency. The magnitude of the shift is a function of the individual design and the change in ambient temperature.

25.2.2.5 Tunability. The basic magnetron design has been utilized in systems with various tuner actuators. Ultimate rates of tuning are limited by inertia in the tuning elements themselves or in the drive mechanism (deformation of the parts results from excessive force). When direct reversing drive systems are used motor inertia is the limiting factor. When a differential shaft driven by two oppositely rotating motors is used to actuate a screw thread, and the servo amplifier has anticipation circuits, almost any characteristic can be obtained. The most successful system to date utilizes a servo-actuated valve controlling flow to a hydraulic tuner motor. Performance of existing systems and the ultimate capability of the various mechanisms are readily compared in Table 25-4.

TABLE 25-4. Tunability Characteristics

Tuner type	Present Rate		Feasible Rate	
Screw threads, driven by reversible motor	100	(Mc/sec^2)	100	(Mc/sec^2)
Screw threads, driven by differential shaft from two counterrotating motors	250		2,000	
Hydraulic, electronic servo-actuated	5,000		100,000	

25.3 Barratrons*

The self-noise-generating tubes which can be video modulated to enhance their already wide rf and video spectrum has been announced, and initial field tests are now underway. This crossed-field device consists of a magnetronlike anode surrounding an asymmetric cathode. To obtain wide bandwidth, it is very heavily loaded.

The Barratron* may be usable with minor refit of the apparatus in those systems now using magnetrons. The tuning range currently available in Barratrons* is approximately half that obtained from a magnetron operating at the same frequency. Due to the very low Q of the anode, a very good load isolator must be used if long-lines effects (Reference 7) (holes) are to be avoided.

*Trade mark registered.

Measurements of performance are currently being made. Table 25-2 summarizes current data. Self-modulation bandwidth is about 2 percent, and external video modulation can increase this to 5 percent. Figure 25-8 shows the spectrum of an S-band tube. Very low frequency and very high frequency video has been measured in the AM detected output of the tube. Figure 25-9 shows the output of a 2-cycle bandwidth receiver tuned to the 10-cycle video output. Figure 25-10 shows the video output of a 1-Mc bandwidth receiver, and Figure 25-11 shows how the density of the photo varies with amplitude of the signal. These all have very noiselike characteristics. Scanty field-test data show that the high peak-to-average power indicated by

(A)   (B)
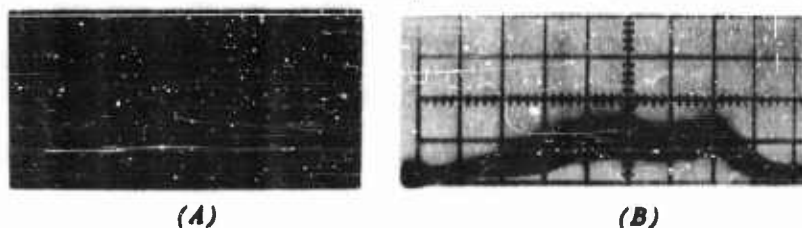
FIG. 25-8. S-band Barratron (trade mark registered). (a) Self-noise spectrum. Scale: 50 Mc per division. (b) Video-modulated spectrum.

FIG. 25-9. Narrowband (2 cps) study of the video output at 10 cps, showing statistical character.

FIG. 25-10. Detected video output from a 1-Mc portion of the Barratron in Figure 25-8a.

Figure 25-11 has been able to break lock on tracking radars at surprisingly low values of power in the receiver bandwidth.
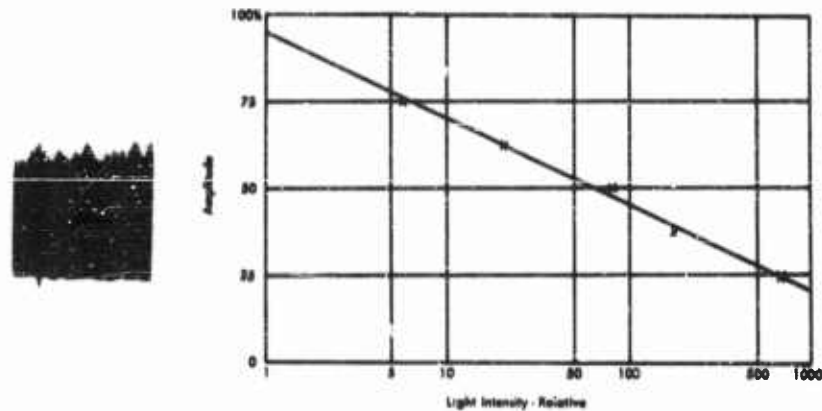
FIG. 25-11. Density of oscilloscope light output vs. vertical position through the trace.

25.3.1 Future Possibilities of Mechanically Tuned Crossed-Field Oscillators

While the advent of the Barratron* may extend the era of their utility, the long-range future of the mechanically tuned crossed-field oscillators is not bright. Tuning rates can probably be increased, but cannot hope to keep up with pulse-to-pulse frequency-agile diversity radars. The only hope is that honest barrage capability can be achieved by placing a number of noise spectra of Barratrons* side by side, thus barraging a wide frequency band. The high efficiency and light weight of these devices promise to continue their usefulness indefinitely.

25.4 Klystrons

The microwave tube requirements of countermeasures systems have led to the development of some new types of klystrons and the application of more standard types. While the present trend appears to be toward voltage-tunable devices for transmitters and for local oscillators of receivers, the

*Trade mark registered.

klystron nevertheless has some characteristics which have been useful for both of these functions.

25.4.1   Characteristics of Klystrons

As a microwave transmitter the klystron is distinguished as having the highest power capability of any type tube. Other characteristics, such as gain, tuning, modulation, efficiency, and load sensitivity, are pertinent to the various kinds of tubes. In actual countermeasures uses, the emphasis has been on oscillators, particularly the "floating drift tube" klystron.

25.4.2   Principles of Klystron Operation

The theory of the klystron has been developed with a good degree of completeness for a microwave tube, and an extensive literature available. For the reader interested in using the tubes, the treatments in books such as Harrison (Reference 8), Beck (Reference 9), Warneke and Guenard (Reference 10), and Spangenberg (Reference 11) give a good outline of the subject, and references to original sources. A brief elementary discussion is given here as an introduction to the particular types of klystrons useful in countermeasures.

Basically, the klystron is an energy transducer operating by velocity modulation of an electron beam. It is distinguished from other beam devices by the use of resonant cavities as circuit elements to couple energy into and out of the beam. To illustrate the action, Figure 25-12 shows the simplest klystron, a basic two-cavity amplifier.

In most klystrons, the electrons move in a roughly cylindrical beam, in the axial direction. The beam originates from a thermionic cathode, typically the concave surface of a spherical cup. The electrons are drawn toward an anode which is shaped to create a radial electric field, similar to that between two concentric spheres. In this way, the electrons converge along radii of the cathode sphere to form a beam smaller in diameter than the emitting cathode. By making this "electron gun" highly convergent, the emission density required of the cathode for a given beam current is reduced, and the life prolonged.

In low-powered tubes, the anode is sometimes made as a grid which the beam goes through. In high-powered tubes a grid would be melted by the energy of electrons intercepted on the mesh, so the anode simply contains a hole through which the beam penetrates into the rf section of the klystron.

The cathode-anode region of the klystron constitutes a spherical diode in which the current flow is governed by the familiar space-charge conditions.

$$I_o = kV_o{}^{3/2}$$

Fig. 25-12.   Schematic klystron amplifier. The basic two-cavity amplifier is the simplest klystron. Gridless cavities and magnetically focused cylindrical beam are typical of power klystrons.

where I_o is the beam current, V_o is the beam voltage, and the constant k, a function of the geometrical arrangement of electrodes, is the perveance. A theoretical solution for the potential and current flow in a spherical diode was derived by Langmuir and Blodgett (Reference 12), giving the perveance as a function of the radii of curvature of the anode and the cathode.

In a practical electron gun, only a conical sector of the sphere can be used. Outside this sector there is no current, so the space potential would be different from that inside the beam where space-charge is present. Pierce (References 13 and 14) has shown how focus electrodes beyond the beam edges may be shaped to produce a potential distribution matching that of the space charge at the beam edge so that all the electrons in the beam see only radial electric field and follow proper radial paths. For a spherical sector, the perveance is then the area fraction of that of a complete sphere.

In high-power gridless klystrons, perveance values are usually in the range from 0.5×10^{-6} to 2.0×10^{-6} amp/volt$^{3/2}$. In low-power tubes where the hole in the anode is covered with a grid, the anode can be much closer to the cathode, and perveances as high as 10×10^{-6} may be used.

If it is desired to amplitude modulate the klystron output, the diode electron gun may be converted to a triode by adding a control grid as shown in Figure 25-12. The grid is also spherical, and controls the emission as in any triode. To avoid distortion of the electron optics, the grid should have a fine mesh, hence the triode has a high amplification factor and the grid must run positive with respect to the cathode.

The electron beam, on entering the hole in the anode, enters a region in which the only electrostatic field is the space-charge repulsion between electrons. The beam tends to spread out. However, the beam enters this region with a converging motion from the spherical cathode, so the resultant shape is convergent down to a minimum diameter point, after which it diverges again; and if no other forces are introduced it will expand indefinitely.

The power klystron illustrated in Figure 25-12 uses a magnetic field to keep the beam focused in a cylindrical outline. At the point where it reaches minimum diameter under the influence of the electrostatic and space-charge forces, it enters a magnetic field directed along the beam axis. This field terminates on an iron pole piece, the beam entering through a hole in the pole piece. In the hole, the magnetic flux lines diverge radially to enter the iron, and all electrons which are not on the axis cross the flux lines and experience a force causing them to rotate about the axis. Proceeding into the region of uniform field, the rotational velocity cutting the axial field produces an inward force. There is a particular value of magnetic field for which the force exactly balances the combination of mechanical centrifugal force and space charge repulsion. Then the beam stays at the same diameter in a condition known as "Brillouin Flow," described by the equation:

$$B = 8.31 \times 10^{-4} I_0^{1/2} V_0^{-1/4} a^{-1}$$

where B is the magnetic flux density in webers/m^2 and a is the beam radius in meters.

Interaction of the electron beam with the rf signal occurs as the beam traverses a gap in the metal tube surrounding it. The first gap crossed is a "buncher" gap. In it, an rf voltage across the gap alternately speeds up and slows down the electrons as the field alternates in direction. The transit time across the gap is considerably less than an rf cycle, so that during the time that any one electron takes to cross the gap, the field does not change much. The rf voltage across the gap is small compared to the beam voltage,

and under these conditions the additional kinetic energy imparted to electrons which are accelerated in the gap is equalized by energy absorbed from decelerated electrons. Thus to a first approximation, no rf energy is absorbed from the cavity in creating the velocity modulation. A second-order effect due to finite transit time does cause some energy to be used in the bunching process.

The input signal to the klystron is fed into the buncher cavity, usually through a coupling iris. The resonant cavity acts as an impedance transformer, matching the signal from a low impedance source, such as a transmission line, to the beam. As pointed out above, the resistive loading on the cavity due to bunching the beam is small, so the impedance step-up or loaded Q of the cavity is high.

The mechanism of amplification in the klystron is by the conversion of velocity modulation to current modulation in the beam. As the beam traverses the buncher gap, it acquires velocity modulation. Then it enters a field-free hollow tube, known as a "drift tube." In this region, the faster electrons catch up with the slower ones. Consider a reference electron which crosses the buncher gap at an instant when the rf voltage is zero and changing from decelerating to accelerating. Another electron crossing the gap just before the reference is slowed down so that our reference particle catches it in the drift tube. An electron crossing after the reference is speeded up so it catches up with the rest. It is seen that this "reference" electron forms the center of a bunch. By the same process, an electron crossing the buncher gap when the field is zero but changing from accelerating to decelerating, finds its neighbors drifting away from it and becomes the center of a region of the beam where the charge density is low. An observer traveling down the beam with the average electron velocity would see the bunches becoming more dense, reaching a maximum concentration, and then dispersing. As viewed from a stationary point in the drift tube, the passing bunches of electronic charge represent an alternating current component superposed on the average d-c beam current.

Mathematically, the a-c component of beam current is given by:

$$I_1 = 2I_o J_1(\theta V_1/V_o)$$

where I_1 is the amplitude of the fundamental component of rf current

θ is the time, in radians of the drive frequency, of drifting at the average beam velocity

V_1 is the amplitude of the input rf voltage producing velocity modulation

This current reaches a maximum value of $1.16I_o$ when $\theta V_1/V_o = 1.84$, giving an rms current of $0.82I_o$.

At the point along the drift tube where the rf current is maximum, the beam crosses the gap of another resonant cavity, the output cavity. Since the lines of force from the electronic charges must end in induced charges on the drift-tube walls, the induced charge must be transferred from one side of the gap to the other as the electrons cross. This motion of the induced charge constitutes a driving current in the resonant cavity. The cavity is coupled to the useful load through an output transmission line, the coupling being adjusted so that the resonant impedance of the cavity matches the source impedance of the electron beam for optimum energy transfer. The phase of the cavity voltage adjusts itself so that the bunches of electrons cross the gap when the electric field is in a direction to decelerate the electrons. Thus energy is transferred from kinetic energy of electrons into circuit energy. The average velocity of the beam electrons is thereby decreased.

After passing through the output cavity, the electron beam is of no further use. It emerges from the focusing magnetic field and is spread apart by the space-charge repulsion between electrons. The beam enters a hollow "collector" and is intercepted on the walls. The residual kinetic energy converted to heat on the collector walls represents most of the power dissipation in the klystron.

At this point, it may be well to point out some reasons for the high power capabilities of klystrons. If the tube operates ideally, no electrons strike any part of the structure except the collector. The rf circuits, whose size is determined by the frequency, do not have to dissipate any power except the circulating current losses. Since the collector is not part of the rf circuits, it can be made as large as needed.

In the same vein, the klystron cathode is also not a part of the rf circuit. This means there is no extraneous heating by electron bombardment or circulating currents. Also, since the electron beam converges after leaving the cathode, the size of the cathode and hence its emission capabilities are limited only by the cleverness of the designer.

25.4.3   The Floating Drift Tube Oscillator

For the generation of large amounts of rf power in jamming transmitters, tunable oscillators have offered a fairly convenient method. There are several disadvantages to a free-running oscillator, such as frequency pulling by mismatched antennas and pushing by voltage variations. On the other hand, the ability to set frequency rapidly by tuning a single circuit makes the oscillator attractive.

The klystron amplifier described above may be converted into an oscillator by feeding part of the output power back into the input cavity in the proper phase. The result is a "two-cavity" oscillator, which may be tuned by tuning

the individual cavities. To simplify tuning, the two cavities may be combined into a single cavity with two rf interaction gaps. The floating drift tube oscillator (Reference 15), illustrated in Figure 25-13, is the type which has
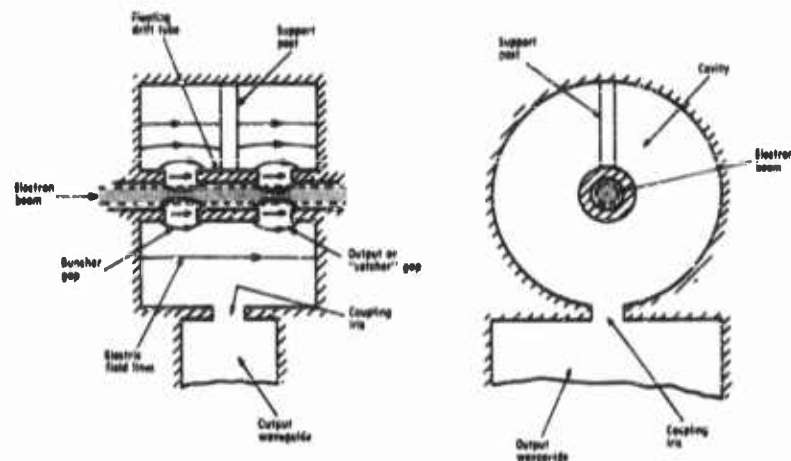
FIG. 25-13. Floating drift tube oscillator. A single cavity with two gaps give strong feedback for the oscillator. The resonant mode illustrated was used in ECM oscillators.

been most used in countermeasures. Here the electric field in the cavity is concentrated in the two gaps by a conducting drift tube between them. The cavity is excited in its fundamental mode, in which the electric field in the two gaps is in phase and in the same direction. Ideally, the drift tube could really "float" with no electrical connection to the outside cavity. In practical tubes, it is supported by an arm from the cavity wall—the arm being positioned on the electric field neutral plane of the cavity so that circulating currents along it are minimized.

The electronic interaction in the floating drift tube oscillator is exactly the same as that of the klystron amplifier. However, the fact that both gaps are part of the same resonant cavity adds the additional restraint that the gap fields are in phase.

As described above, the bunch of electrons is formed about a reference electron which crosses the first gap when the field is zero and becoming accelerating. To convert the beam energy in the output gap, the bunch of electrons should cross the gap when the field is at its maximum decelerating value. One can see from this relation that the transit time between gaps should be $\frac{3}{4}$ of a cycle, or $1\frac{3}{4}$, $2\frac{3}{4}$ cycles, etc. In most of the practical tubes which have been built, the $1\frac{3}{4}$ cycle mode is used, since this cor-

responds most closely to the optimum transit angle for a high-power kly-stron.

The floating drift tube oscillator can be amplitude modulated by using a control grid to vary the beam current. Also, frequency modulation can be produced by varying the beam voltage about its optimum value. This causes the electron bunches to cross the output gap before or after the instant of maximum retarding voltage. The induced current has a quadrature component out of phase with the cavity voltage, which pulls the frequency away from its normal resonance. In this respect the frequency modulation action is the same as that in the familiar reflex klystron oscillator. When AM and FM can be generated simultaneously and with independent modulations, the oscillator gains attractive features for jamming transmitters.

In tuning the oscillator, two adjustments are made. The resonant frequency of the cavity is changed by a mechanical motion. Also, since the transit angle between gaps must be constant, and the velocity of the electron beam is proportional to the square root of the voltage, the beam voltage must be varied as the square of the frequency.

25.4.4   Practical Floating Drift Tube Klystrons

25.4.4.1   The V-22, 500-Watt X-Band Oscillator (SECRET). The most highly developed floating drift tube oscillator that has been built is the Varian Associates Model V-22. Figure 25-14 shows a photograph of

FIG. 25-14. Photograph of V-22, 500-watt oscillator using permanent-magnet focusing. SECRET
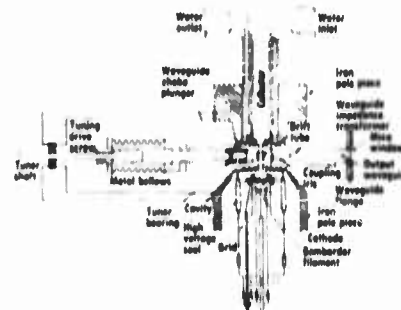
FIG. 25-15. Cross section of V-22, showing important parts of the oscillator. Mechanical details are omitted. C-shaped magnets do not show in this section. SECRET.

the device. This tube was developed under an Air Force prime contract with W. L. Maxson Company as the transmitter in an X-band airborne jammer. For this service it was desirable to cover a wide frequency range with as few tubes as possible without sacrificing performance. Two models of the tube were built, each covering 1.2:1 range with a single tuning control. To give a clearer idea of the practical realization of this kind of a tube, Figure 25-15 shows the essentials of the construction. The cathode is a concave button of thoriated tungsten, heated by being bombarded from the rear by electrons from a crude "primary" electron gun. This method of heating is necessary because the emitting temperature of thoriated tungsten is so high that heating by thermal radiation from a filament will not work—the filament would melt. Around the cathode is a tantalum focusing electrode to define the beam, and directly in front of it is the tungsten mesh control grid.

The beam converges to $\frac{1}{3}$ of its initial diameter and is magnetically focused through the cavity. In this tube, the focus is provided by a permanent magnet, with steel pole pieces which are a part of the vacuum envelope.

The resonant cavity itself contains the "floating" drift tube supported on a copper post through which cooling water circulates. For tuning, the cavity is constructed as a section of waveguide of weird shape. One end is short-circuited and the other terminates in a choke-type plunger. The entire cavity, including sliding bearings for plunger alignment, is evacuated. For tuning, the plunger is moved by deforming a flexible metal bellows.

The collector is an enlarged hollow "bucket" surrounded by cooling water channels. In this high-power device, it is necessary to cool practically everything, including the tuning plungers.

The output power is coupled, through an inductive iris in the cavity wall, into the waveguide. In the waveguide is a frequency-sensitive impedance transformer. This is designed to transform a flat line into the proper load for the oscillator at any frequency in the band. In this case the load varies considerably. As pointed out above, the beam voltage must vary as the square of the frequency. It is generally desirable to maintain constant power, so the beam current is made inversely proportional to the voltage.

$$V_o \sim f^2$$
$$I_o V_o = \text{constant}$$
$$\therefore I_o \sim f^{-2}$$

The impedance of the beam

$$V_o/I_o \sim f^4$$

and the effective load resistance presented to the beam should be proportional to this beam impedance for optimum energy transfer. Transformation of the

matched waveguide impedance to get the variation with frequency can become difficult when large tuning ranges are involved. In the V-22 it is done with a frequency-sensitive transformer consisting of a section of low-impedance waveguide adjacent to the coupling iris.

The output window has always been one of the critical items of high-power microwave tubes. The V-22 was designed before ceramic-to-metal sealing techniques had reached their present state of reliability, so the output window is a thin sheet of mica, sealed across the waveguide with a low-melting glass as sealing cement. The dielectric losses in the mica are quite small, and the glass is in a recessed groove where it is not exposed to very high electric fields, so the mica windows have been successful in operation.

Operating Characteristics of the V-22. The jamming transmitter for the V-22 used an amplitude-modulated signal. The modulation was white noise with a bandwidth of 10 Mc. This was applied to the klystron grid as an alternating voltage with time-average of zero, which was superposed on a d-c bias. The resulting beam-current modulation governs the rf output of the oscillator, as illustrated in Figure 25-16. The foot of this curve, or "starting current," represents the value where the nega-

Fig. 25-16. Amplitude-modulation characteristic, V-22. The relation of oscillator output to instantaneous beam current is a function of the load conductance. Curves are labeled with load conductance values.
SECRET.

tive conductance of the electron beam (for small signals) exceeds the external load conductance as transformed by the cavity. The starting current is thus a function of the load VSWR. Now in applying the noise modulation it is highly desirable that the negative modulation peaks do not drive the beam current below "starting current." If this happens, the signal contains a base-line upon which a radar pulse, for example, is superposed and can be read. This being the case, the peak modulation amplitude was chosen small enough so that the oscillator was not driven below starting current for the worst predicted load match. (Recent developments in high-power isolators can remove this variable and permit better performance.)

A condensation of the important characteristics of the V-22 follows:

Power Output (average)	500 watts
Power Output (modulation peaks)	1000 watts
Beam Voltage	7–10 kv
Beam Current (average)	400 ma
Grid Bias	+25 volts
Grid Modulation Voltage	±25 volts peak-to-peak
Bandwidth	20 Mc
Tuning Ranges	7.5–9.1 kMc
	9.1–11.0 kMc

Figure 25-17 shows typical variation with frequency of the power output, beam voltage, and tuner setting for the low-frequency V-22.
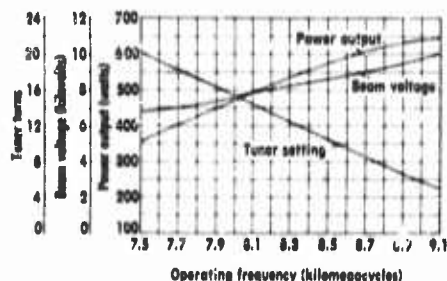
Fig. 25-17. Power vs. frequency, V-22. Beam current is held constant at 0.40 amp. Beam voltage is optimum value, proportional to f^2. SECRET.
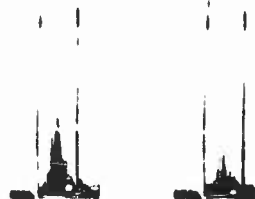
Fig. 25-18. Spectrum, V-22. Noise spectrum obtained by 10-Mc white noise on the grid. Frequency markers are ±10 Mc from the carrier. SECRET

The noise spectrum produced by beam-current modulation turns out to be not pure AM, but contains some frequency pushing also. Figure 25-18 shows output spectra at the low and high ends of the tuning range, showing carrier suppression. The marker lines are at ±10 Mc from the carrier.

This oscillator was mechanically tuned by a motor drive with a servo circuit for frequency setting. The speed was such that the transmitter could be tuned across its range in 10 seconds. The voltage control potentiometers for beam voltage and grid bias were ganged to the main tuning control for automatic operation.

25.4.4.2 The V-71, 5 KW, X-Band Oscillator. A higher-powered floating drift tube klystron was developed in 1954 by Varian Associates for the Evans Signal Laboratory, and was intended for ground-based jamming. This was the V-71, rated at 5 kw CW output. A photograph of the

klystron is shown in Figure 25-19. The high-power tube did not reach the product refinement stage of the 500-watt V-22, but since it introduced some
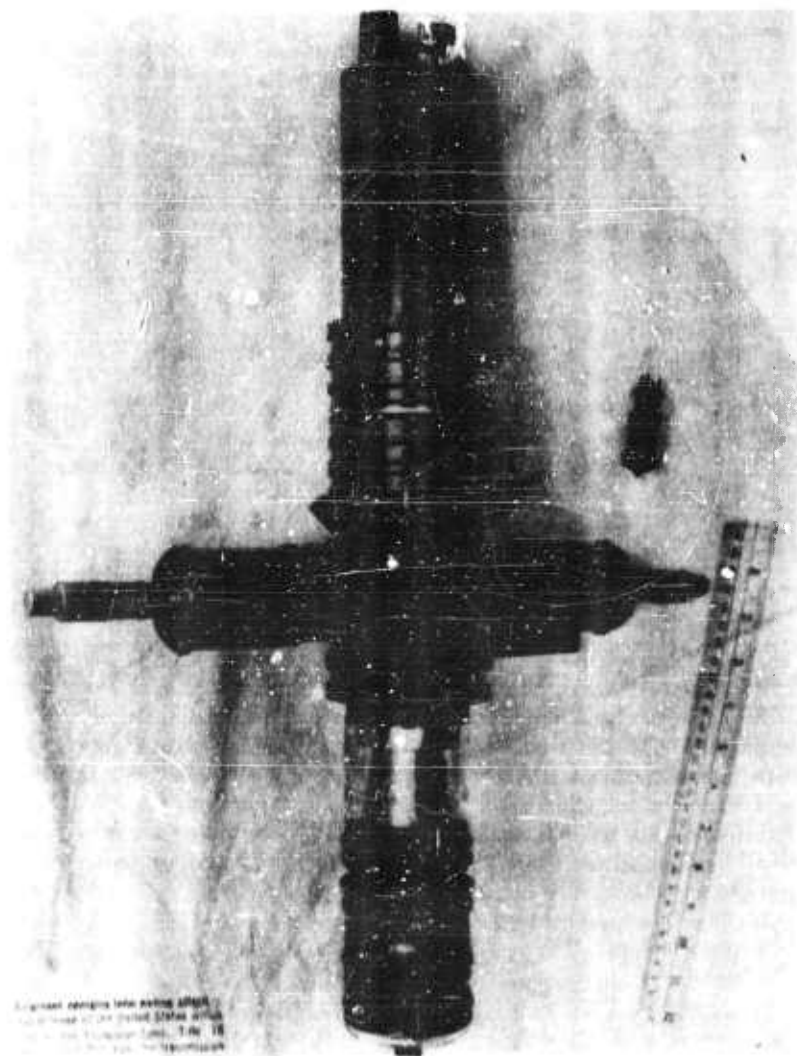
FIG. 25-19. Photograph of V-71. This 5-kw tube used an electromagnet (not shown).
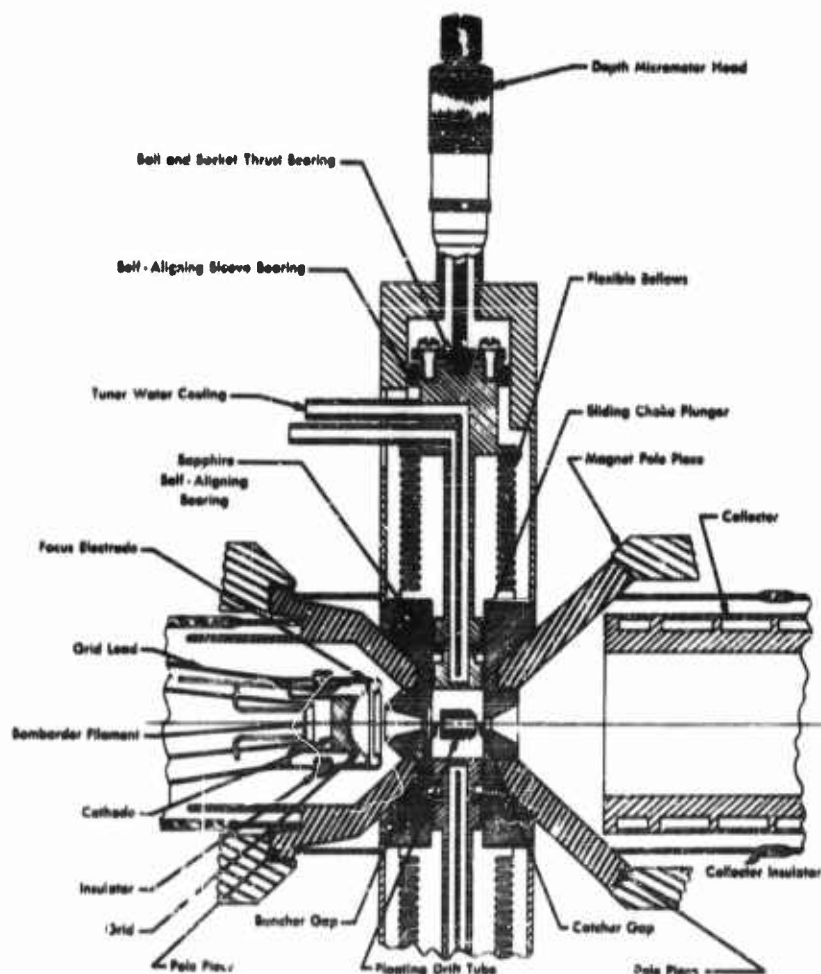
FIG. 25-20. Heart of the V-71. A simplified mechanical layout of the 5-kw oscillator.

new features and represented the farthest development of the klystron power oscillator, it is worth describing.

Figure 25-20 shows a simplified cross-section of the heart of the tube. It is in many respects a "big brother" version of the V-22, but with some important differences. It covers the X-band from 7.5 to 10.5 kMc in one tube instead of two. To achieve this tuning range, it was necessary to use two

choke plungers, one at each end of the waveguide cavity. In this way the
two shorts are moved in and out together, so that the gaps are always at the
central high-impedance point of the cavity. The experimental tubes were
built with independent tuner controls, but since exact synchronism is not
necessary they could easily be ganged.

As explained in the previous section, the velocity of the electrons between
gaps must be varied proportional to the frequency as the tube is tuned.
With a 1.4:1 frequency range, the effective beam voltage variation has to be
2:1 to maintain the same transit angle. If this were done directly, with the
inverse variation in beam current to keep constant power, the d-c impedance
would vary by 4:1. This difficulty was eliminated in the V-71 by having the
drift tube insulated for d-c from the rest of the tube body. The drift tube is
mechanically supported on an insulated post. The capacities of the two gaps
in the resonant cavity are made equal so that no net current is induced in
the support post. Also, the post enters the cavity through an rf choke section
for isolation. A bias voltage (negative to ground) is applied to the "floating"
drift tube to regulate the effective potential within it (with respect to the
cathode), which determines the electron velocity. Thus the main power-sup-
ply voltage and current from cathode to ground are not varied with fre-
quency. A diagram of the power supply and modulation inputs to the klystron
is shown in Figure 25-21.
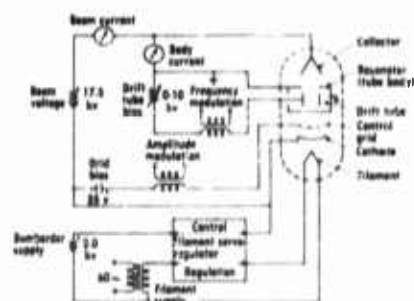
Fig. 25-21. Power supplies for floating
drift tube oscillator. Amplitude and fre-
quency modulation are applied independ-
ently. Transit angle is controlled by the
drift-tube bias voltage.

By applying a video modulating
voltage to the drift tube, the electron
velocity and hence the phase angle of
the rf current in the output gap is
modulated. The effect is exactly analo-
gous to modulating the beam voltage
in a conventional floating drift tube
oscillator or the reflector voltage in a
reflex klystron—the varying phase an-
gle of the rf beam current pushes the
oscillator frequency away from the
normal cavity resonance and produces
frequency modulation. Since the drift
tube draws very little current, the
dynamic modulation impedance is high,
making it easy to apply FM to the
oscillator.

Amplitude modulation is produced by modulating the beam current with
the control grid. Since the amplitude and frequency modulations are largely
independent, various combinations may be used.

The operating characteristics of the V-71 are:

Power Output (CW)	5 kw
Tuning Range	7.5–10.5 kMc
Beam Voltage	17.5 kv
Beam Current	3.0 amp
Drift Tube Bias	10.0–3.0 kv
Bandwidth	20 Mc
Modulation Sensitivity (FM)	7 Mc/kv
Grid Modulation Voltage (AM)	±20 volts peak-to-peak

25.4.5 Evaluation of Klystrons

The career of klystron oscillator jammers was short, starting from the need for high-power high-frequency tunable transmitters and ending with frequency agility radar techniques and voltage-tunable jammers.

Possible future development might lie in the very high power field, where the klystron amplifier has great capability. At power levels above 1 Mw the modern stagger-tuned klystron rivals the traveling-wave amplifiers in electronic bandwidth (12 percent achieved in one case). Saturated efficiency of 40 percent is typical. With the stability and high gain of the klystron amplifier, the over-all system efficiency can approach this number.

REFERENCES

1. George B. Collins, *Microwave Magnetrons*, McGraw-Hill, New York, 1948.

2. L. Marton, *Advances in Electronics and Electron Physics*, Academic Press, New York, 1956. See section by E. C. Okress, "Magnetron Mode Transitions," with particular reference to the extensive bibliography on page 531.

3. J. B. Fisk, H. D. Hagstrum, and P. L. Hartman, "The Magnetron as a Generator of Centimeter Waves," *Bell System Tech.*, Vol. 25, pp. 167-348, 1946.

4. A. W. Hull, *Phys. Rev.*, Vol. 18, p. 31, 1921.

5. D. R. Hartree, CVD Report.

6. D. A. Wilbur, General Electric Research Laboratory reports under U. S. Army Signal Corps Contract W-36-039 sc-32279.

7. J. F. Hull, G. Novick, and R. Cordray, "How Long Line Effect Impairs Tunable Radar," *Electronics*, February, 1955.

8. A. E. Harrison, *Klystron Tubes*, McGraw-Hill, New York, 1947.

9. A. H. W. Beck, *Velocity-Modulated Thermionic Valves*, MacMillen, New York, 1948.

10. R. Warneke and O. Guenard, *Les Tubes electroniques à commande par modulation de vitesse*, Gauthier-Villars, 1951.

11. Karl R. Spangenberg, *Vacuum Tubes*, McGraw-Hill, New York, 1948.

12. I. Langmuir and K. Blodgett, "Currents Limited by Space Charge between Con-
    centric Spheres, *Phys. Rev.*, Vol. 24, pp. 49-59, 1924.

13. J. R. Pierce, "Rectilinear Electron Flow in Beams," *J. Appl. Phys.*, Vol. 11, pp. 549-
    554, 1940.

14. J. R. Pierce, Theory and Design of Electron Beams," Van Nostrand, Princeton,
    New Jersey, 1949.

15. M. Chodorow and I. P. Fan, "A Floating-Drift-Tube Klystron," *Proc. IRE*, Vol.
    41, p. 25, 1953.

26

O-Type Microwave Tubes

D. A. WATKINS,   C. B. CRUMLY

26.1 Introduction

The practical development of microwave technology was made possible by the invention and development of the magnetron and klystron tubes, which made available large amounts of power in the microwave range, that is, from about 1,000 to 30,000 megacycles per second. With the advent of traveling-wave tubes, the two chief shortcomings of the former tubes, namely, their narrow bandwidth and slow tuning capabilities, were largely overcome. Thus, the door was opened to a new class of applications in which operation over a very broad band of frequencies, and essentially instantaneous tuning, are important properties.

The field of electronic countermeasures is perhaps the best example of a class of applications in which these unique features of traveling-wave devices can be put to good use. An amplifier which can simultaneously receive signals within a two-to-one band of the spectrum makes possible a new class of receiving systems. An oscillator capable of changing its frequency over a wide range in a few microseconds can be the basis of a sophisticated jamming system. It is because of these two inherent properties of traveling-wave devices, broad bandwidth and voltage tuning, that their applicability in electronic countermeasures systems is so firmly established. In this chapter and the next, the basic kinds of traveling-wave devices, their principles of operation, and their important characteristics will be discussed.

Inasmuch as it would be virtually impossible to give complete acknowledg-

ment to the many workers responsible for the development of the technology in this chapter, a bibliography of typical and informative material has been prepared, and appears at the end. It has been arranged by subjects, in order to facilitate the search for more detailed information on a particular matter of interest to the reader.

26.2 Traveling-Wave Devices

Traveling-wave devices make use of a gain or interaction mechanism between an electron beam and *traveling* electromagnetic wave. In contrast, klystrons and magnetrons depend upon *standing* waves to produce the high-frequency fields with which the beam interacts. Further, the interaction process in a traveling-wave device is *distributed* in space, rather than being localized as in the klystron. Since standing waves imply repeated reflections back and forth on a microwave circuit, strong fields require resonant circuits, and standing-wave devices are inherently narrow in bandwidth. On the other hand, traveling waves can be produced on nonresonant, broadband circuits, and consequently, traveling-wave devices can operate over extremely broad bandwidths. Since the fields produced by nonresonant circuits are generally much smaller than those of resonant circuits, a stronger interaction process is required to give a similar effect. This condition is satisfied by the distributed nature of traveling-wave interaction, in which the gain mechanism is continuous and cumulative throughout the entire region of the circuit.

Traveling-wave devices are classified according to the nature of the interaction process between the beam and the wave. In O-Type tubes, the effective r-f electric field produced by the traveling wave is parallel to the direction of motion of the beam of electrons, and produces a slowing-down of the average velocity of the d-c beam. The wave is amplified by the exchange of energy from the beam to the wave, and this energy comes from the initial d-c *kinetic* energy of the electrons. Generally, the beam is constrained to move in a straight line along the axis of the r-f structure by a strong magnetic field or other focusing arrangement. There is no d-c electric field present in the interaction space.

M-Type tubes, or crossed-field devices, operate on a different interaction process which depends, not on a slowing-down of the beam's motion in the direction of wave propagation but upon abstraction of energy from a transverse d-c electric field by transverse motion of the electrons. The r-f field produced by the traveling wave has components both parallel and perpendicular to the motion of the beam which are effective; in addition a d-c electric field is produced transversely, across the beam, by suitable electrodes. The r-f field, instead of slowing down the beam, causes it to move in

a sidewise direction toward the more-positive electrode. The average speed of the beam remains essentially unchanged; the *potential* energy of the beam is converted into r-f energy of the wave, which is therefore amplified. In order to keep the beam moving along the axis of the tube, despite the strong transverse d-c electric field, a magnetic field, perpendicular to both the tube axis and the electric field, is required.

(It should be noted that, while the above description applies to a *linear* M-Type tube, it makes no essential difference whether the "tube axis" is a straight line, or is wrapped up into a circle.)

The "voltage-tuned magnetron" is usually classed as a third type of microwave tube, although the mechanism of operation is essentially that of an M-Type tube. The VTM is an oscillator formed by wrapping an M-Type tube into a circle, and closing both the circuit and the beam upon themselves. The resulting device is an oscillator capable of voltage-tuning over an extremely wide range in frequency.

A summary of the classifications of microwave traveling-wave devices is
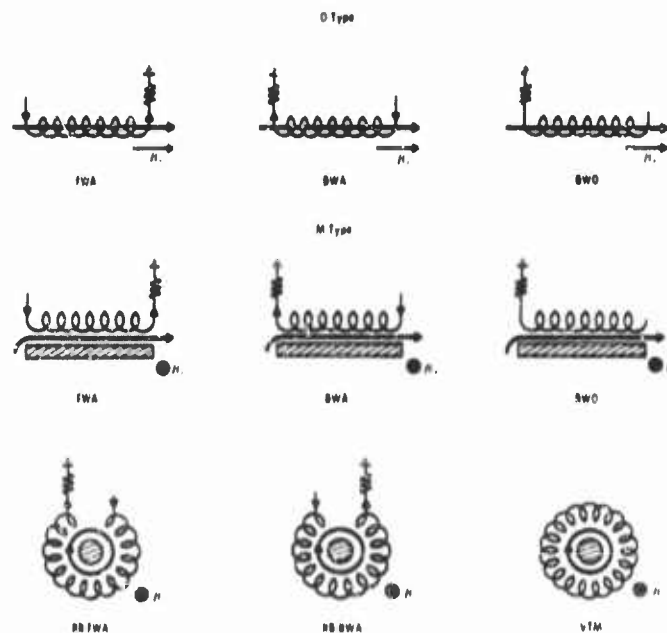
FIGURE 26-1   Schematic drawings of O-Type and M-Type tubes

given in Figure 26-1. This chapter will be concerned with the three kinds of O-Type tubes shown: (1) the forward-wave amplifier (FWA) in which the

r-f power flow on the helix or other circuit is in the same direction as the flow of electrons; (2) the backward-wave amplifier (BWA) in which the direction of r-f power flow is reversed relative to the beam travel; (3) the back ard-wave oscillator (BWO) which is the same as the BWA, except operated at high enough current so that an output is obtained without any input. The M-Type tubes will be taken up in the next chapter.

26.3 The Traveling-Wave Amplifier

The traveling-wave tube is a device which produces amplification of a signal traveling along a circuit, by means of distributed interaction with a beam of electrons moving parallel to the circuit. The essential features of such a tube are illustrated in Figure 26-2, and consist of a slow-wave cir-
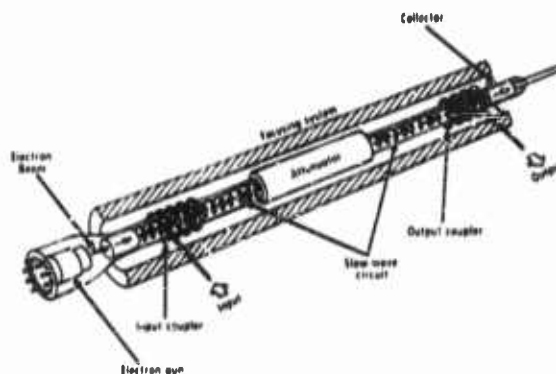
FIGURE 26-2   A traveling-wave amplifier, showing the essential elements

cuit, an electron gun and focusing system, and an attenuator, together with transducers for coupling the signal into and out of the circuit.

The slow-wave structure functions to retard the phase velocity of the signal wave to a value substantially less than the velocity of light, so that approximate synchronism can be obtained between the signal wave and the electron beam, for reasonable beam accelerating voltages. In addition, the circuit must be such that the signal wave produces an r-f electric field parallel to the direction of beam motion, at least in the region occupied by the beam.

Under the proper conditions of near synchronism, the electrons then interact continuously with the wave over a considerable distance, resulting in a conversion of the beam kinetic energy into signal energy on the circuit. It is thus possible to obtain very high amplification of the signal wave. Further, since the slow-wave structure can be inherently a nonresonant circuit, in

which the phase velocity is relatively independent of the signal frequency, this amplification can be obtained over an extremely broad band of frequencies, with no mechanical or electrical adjustments whatever.

26.3.1 Mechanism of Operation

For a *helical* slow-wave structure, if the frequency is not too low, the r-f wave travels essentially along the wire at the velocity of light, so that the component of its velocity along the axis of the helix is reduced by a factor closely approximating the ratio of helix pitch to helix circumference. Further, if the frequency is not too high, this wave produces a strong electric field on the axis of the helix, which is parallel to this axis. Thus, if a beam of electrons is injected along the axis of the helix, with a velocity which is nearly synchronous with that of the axial field, the electrons will be subjected to slowly varying accelerating and decelerating forces. An interaction takes place between the electrons and the field, which, on the average, can cause the electrons to give up energy to the field.

This interaction process, which results in amplification of the signal wave, can be understood by reference to Figure 26-3. Here it is assumed that a

signal wave of constant amplitude is impressed on the helix, which results in a distribution of electric field on the axis as indicated by the curve in Figure 26-3a. (If the observer moves along with the wave, the field distribution shown will remain stationary.) The polarities are so chosen that a *positive* field *accelerates* the electrons; and if the direction of travel is taken to the right, then a positive field produces motion to the right, negative to the left. Consider that, near

FIGURE 26-3 Diagram illustrating the bunching process in a traveling-wave amplifier

the input end of the tube, a uniform distribution of electrons has been injected into this field. Electrons which are at positions labeled 2 and 5 are in zero field, and hence tend to remain where they are. Electrons 3 and 4 find themselves in an accelerating field, and thus begin to move to the right, toward position 5. Electrons 1 and 6 are in a decelerating field, and begin to move to the left. The effect of the electric field is then to produce a velocity modulation on the beam, in such a manner that bunches of electrons begin to form at positions such as 5. Similarly, an "anti-bunch", or a

deficiency of electrons, will form at positions such as 2. This velocity modulation and bunching process continues to take place along the entire tube, and is a cumulative process. That is, the longer the beam is subjected to the field, the tighter the bunches grow.

As the circuit wave is affecting the beam as described above, the bunched beam is simultaneously affecting the circuit. A bunched beam, traveling near a circuit of this type, will induce a wave on the circuit which, in turn, will produce an electric field on the axis. The phase of this induced wave is such that the position of the bunch coincides with the point of maximum decelerating force, as indicated in Figure 26-3c. The induced wave then lags the original signal wave by a quarter wavelength. This induced wave is combined with the signal wave to produce the resultant wave in Figure 26-3d, which has moved back in phase somewhat, with respect to the bunch. The electrons in the bunch therefore see a retarding field, and are slowed down, delivering energy to the wave on the helix. The velocity modulation and bunching processes continue to take place, even as the bunched beam continues to convert its kinetic energy to circuit-wave energy, and hence the wave on the circuit grows in amplitude as beam and wave move down the tube. The bunching becomes more complete; since the induced wave grows in amplitude, the resultant wave falls farther behind the bunch, and the bunch therefore sees a stronger retarding field. Thus, a large and increasing amount of energy is delivered to the wave on the helix, which rapidly becomes much larger than the original signal. Analysis shows that the wave on the helix eventually increases exponentially with distance, and large power amplifications can easily be attained, 30 to 60 decibels being typical.

26.3.2 Slow-Wave Structures

The slow-wave structure for an O-Type traveling-wave tube must satisfy two basic requirements. First, it must produce an axial electric field at the position of the beam; and second, the phase velocity of the wave producing this field must be smaller than the velocity of light. There exist many classes of structures which satisfy these requirements, and are thus potential circuits for traveling-wave devices.

The most common structure used in amplifiers is the helix, which not only satisfies the above requirements, but does so over an enormous range of frequencies. Thus, helix-type amplifiers can typically be built to amplify over a two-to-one range with ±3 decibel variation in gain, and even higher ratios are practicable.

The most serious limitation of the helix is its relatively low power-handling capacity. A slow-wave circuit must be able to withstand both the heat generated by ohmic losses due to the signal itself, and also the dissipation due

to any beam interception which may occur from imperfect focusing. All attempts to increase the power-dissipating ability of the helix have met with success only at the expense of some other desirable property, generally its bandwidth.
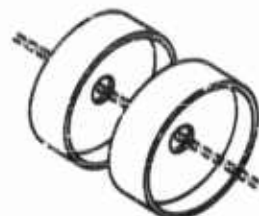
Circular waveguide with apertured discs
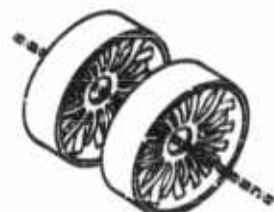
Cavities with interlaced loop coupling

Cloverleaf structure

FIGURE 26-4 Some slow-wave structures that can be used in traveling-wave tubes

Some alternative structures are shown in Figure 26-4. The disc-loaded waveguide is one of the simplest slow-wave structures which has excellent power-handling ability, but it is a dispersive circuit, that is, the wave velocity varies with frequency. Therefore, the frequency range over which near-synchronism between the beam and wave is obtained will be smaller than for a helix, and hence the bandwidth is smaller. The other structures shown are typical of high-power pulsed traveling-wave amplifiers, and represent refinements designed to raise the dissipation ability and yet maintain as wide a bandwidth as possible.

26.3.3 Attenuators

Since the slow-wave circuit can carry a signal in either direction, a traveling-wave amplifier has a built-in feedback path if end matches are not perfect. In order to stabilize such an amplifier, some means must be provided to prevent the large output signal from reaching the input. This is accomplished by loading the slow-wave structure with an attenuator made of some lossy material at a point reasonably close to the input end of the tube. The attenuation of a wave traveling toward the input must exceed the net forward gain of the tube to prevent oscillation; and usually exceeds the gain by 20 to 30 decibels or so, in order to prevent regenerative "wiggles" in the gain versus frequency curve. If the attenuator is bilateral, it will affect the wave traveling toward the output in the same way as the reflected wave, and hence the gain is reduced. However, since the bunched beam is not directly

affected by the attenuator, the reduction in gain is less than 6 decibels, for a section in which the attenuation is infinite. It is therefore possible to obtain substantial gain, with good stability.

In low-power tubes, the attenuator commonly takes the form of a coating of colloidal graphite (Aquadag) brushed or sprayed on the glass envelope or on ceramic rods used to support the helix. In medium-power tubes, coupled-helix attenuators are often wound of lossy wire which are then placed over the outside of the tube envelope. Special impregnated-ceramic materials are being developed for high-power tubes.

The placement and distribution of the attenuator has a critical effect on the efficiency of high-power tubes. It has been found that a relatively short length of attenuation placed near (but not at) the input gives the optimum efficiency. The output section of the circuit should be as nearly loss-free as possible.

It is also important to provide a smooth transition into the attenuator, so that the circuit waves will be absorbed without appreciable reflection. Otherwise, feedback oscillations may occur in the sections of the tube outside the attenuator.

A TWT with a properly-applied attenuator is essentially a stable *unilateral* amplifier, with high *gain* in one direction, and high *loss* in the opposite direction.

26.3.4 Guns and Focusing

The electron guns used in TWTs are generally high-current, low-voltage sources. For this reason, space-charge effects usually determine the limits on what kinds of beams can be produced and maintained. In addition, because of the nature of the interaction process, beams with large ratios of length to cross section are required. Thus, some method of focusing is essential, to prevent the beam from spreading apart under the influence of its space-charge forces.

In early versions, TWTs were immersed in strong axial magnetic fields, which constrained the beam to move along a straight line. In that case, a parallel-flow Pierce-type electron gun could produce the desired beam, as illustrated in Figure 26-5a. The situation where the magnetic field threads the cathode is called "confined flow", since the electrons are confined to straightline paths from emission to collection. The requirement for a strong magnetic field implies the need for a solenoid enclosing the tube and its capsule. Since the weight, bulkiness, and power loss of a solenoid is objectionable in many applications, a good deal of effort has gone into the development of other methods of beam focusing.

A substantial reduction in the magnet requirements can be obtained by

FIGURE 26-5   Pierce-type electron guns

shielding the gun from the field, which causes the beam to spin slightly as it enters the interaction region of the tube. In this condition, the beam is said to be "magnetically focused", in "Brillouin flow". A distinct advantage of this situation is that the gun is no longer in a strong magnetic field, and a converging geometry can be used, as in Figure 26-5b. Thus, the current density that can be obtained in the beam is higher than for confined flow. This focusing method is almost universally used in high-power solid-beam devices.

In low-power tubes, the solenoid can be completely eliminated by the use of permanent magnets or electrostatic fields. The most common scheme in use of TWT's, employs disc-shaped permanent magnets which produce a periodically-alternating axial field. This system, shown in Figure 26-6, has been successfully applied to tubes producing up to

FIGURE 26-6   Periodic-magnetic focusing system

about a watt of output power, roughly from S-band to X-band.

It is possible to produce a uniform field in the region of the beam by the use of a properly-shaped permanent magnet, and reasonably light weights can be attained at S-band and above.

At frequencies below about one kilomegacycle, hollow beams are generally used, for reasons having to do with practical limits on the length of the tube (see Section 26.5). Other than confined flow, which requires very large

Figure 26-5   Pierce-type electron guns

shielding the gun from the field, which causes the beam to spin slightly as it enters the interaction region of the tube. In this condition, the beam is said to be "magnetically focused", in "Brillouin flow". A distinct advantage of this situation is that the gun is no longer in a strong magnetic field, and a converging geometry can be used, as in Figure 26-5b. Thus, the current density that can be obtained in the beam is higher than for confined flow. This focusing method is almost universally used in high-power solid-beam devices.

In low-power tubes, the solenoid can be completely eliminated by the use of permanent magnets or electrostatic fields. The most common scheme in use of TWTs, employs disc-shaped permanent magnets which produce a periodically-alternating axial field. This system, shown in Figure 26-6, has been successfully applied to tubes producing up to

Figure 26-6   Periodic-magnetic focusing system

about a watt of output power, roughly from S-band to X-band.

It is possible to produce a uniform field in the region of the beam by the use of a properly-shaped permanent magnet, and reasonably light weights can be attained at S-band and above.

At frequencies below about one kilomegacycle, hollow beams are generally used, for reasons having to do with practical limits on the length of the tube (see Section 26.5). Other than confined flow, which requires very large

solenoids at these frequencies, the most promising focusing methods turn out to be primarily electrostatic. Development of lightweight focusing systems for hollow beams has not reached a state comparable to that for solid beams, but some of the more promising avenues include: combinations of periodic electrostatic fields and a uniform (radial) field; periodic fields both inside and outside the beam; uniform radial field acting on a spinning beam; and periodic magnetic fields in combination with uniform or periodic electric fields.

The electron guns used in TWTs are generally of the Pierce type, shown in Figure 26-5, nonconverging or convergent, depending on the focusing method used. The perveance of these guns is limited, approaching about 2×10^{-6} for solid beams. Other types are available, such as the Müller gun or the Heil gun, shown in Figure 26-7, when somewhat higher perveances are

Heil type
(1)

Müller type
(ii)

Figure 26-7   High-current electron guns

required. In all cases, the aim is to control the beam trajectory in the gun, by a balancing of space-charge forces, accomplished by beam-forming electrodes.

26.3.5 Bandwidth and Gain

The amplification process in traveling-wave devices is inherently non-resonant, and as a result, amplification can be obtained over very wide bandwidths. The principal factors which cause the gain to vary with frequency are: (1) variation in the velocity of the wave on the circuit with which the beam interacts; (2) variation in the effective length of the tube, in wavelengths; (3) variation in the strength of the axial electric field at the position of the beam; (4) variations of the impedance match at the input and output transducers, and at the attenuator.

The helix circuit (Section 26.3.2) is particularly suited to achieving wideband operation. The wave velocity on a helix is relatively independent of frequency over a wide range. Furthermore, the strength of the axial field decreases in such a way, with increasing frequency, that this factor tends to

compensate for the increasing number of wavelengths along the tube. Thus, the design parameters can be chosen so that the gain versus frequency curve reaches a very broad peak near the center of the operating band. Another important consideration is the ease with which helices can be fabricated. Finally, it is possible to obtain an excellent impedance match to a helix over a large frequency range. A typical curve of gain versus frequency for a helix-type tube is shown in Figure 26-8.
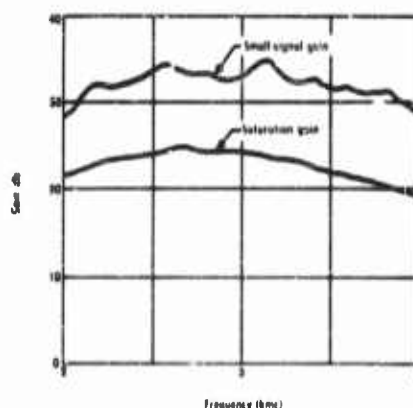
FIGURE 26-8 Typical curve of amplification
of a traveling-wave tube as a function of
frequency

The small-signal gain of a traveling-wave amplifier depends in a complicated way on a number of parameters, including the length of the circuit, the beam voltage and current, the helix interaction impedance, and circuit attenuation. Typically, stable gains of from 30 to 60 decibels can be achieved in a single tube.

26.3.6 Nonlinearities

A typical curve of output power versus input power for a TWT is shown in Figure 26-9. The characteristic is linear up to a point approximately six decibels below maximum output, then begins to saturate, reaches a maximum, and then drops off as the input continues to increase. If operation is well below saturation the tube behaves as a linear amplifier with negligible nonlinear effects. As the input signal level is raised, however, the bunching of the beam becomes nonsinusoidal, and some important nonlinear effects begin to take place.

Limiting. If only one signal is present, the most important nonlinear effect is a compression, or limiting, as indicated by the input-output curve. At the
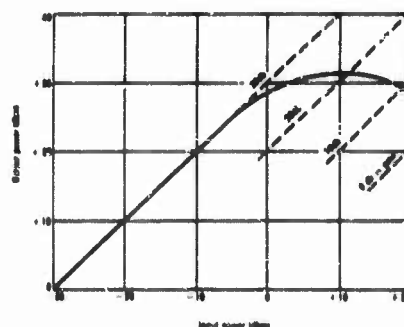
Figure 26-9
Typical saturation curve
of output power
as a function
of input power

point of maximum output, a relatively large variation in the input power (about 10 db) causes only a small variation in the output (about 2 db). If the signal to be amplified contains amplitude modulation, this modulation will be compressed or largely removed when the tube is operated near saturation. At the saturation point, distortion of the modulation signal is large. This limiting property can be emphasized by proper design, if desired, to obtain a much broader saturation curve. A drawback of such traveling-wave limiters is the relatively large amount of phase distortion encountered in the region of saturation.

Harmonics. In the saturation region, the bunched beam contains harmonics of the fundamental signal frequency. If these harmonics lie within the passband of the tube, they can be coupled out, and will appear in the output. Generally, the double-frequency component is the only harmonic of practical importance.

Mixing. When two or more signals are present simultaneously, one of which is strong enough to drive the tube into the saturation region, intermodulation products or beat-frequency components are produced, as in any nonlinear device. The outputs obtained at the beat frequencies depend not only on the strengths of the beating signals and the curvature of the characteristic, but also on the gain of the traveling-wave tube at the beat frequencies. The efficiency of the output transducer also enters in here. It is possible to build traveling-wave mixers which emphasize these effects, and put out essentially full saturated power at a selected beat frequency.

Frequency Conversion. A traveling-wave amplifier can be made to oscillate by the use of internal or external feedback, and the level of the output is determined by the saturation characteristic. If a small signal at another frequency is introduced into the tube, a beat will be formed by mixing with the large oscillation signal, producing an output. This process is not essentially different from mixing, except that the local oscillator signal is supplied by the traveling-wave tube itself.

Signal Suppression. When two large signals are present simultaneously, the nonlinear effects are more complex. The interaction between the two fre-

quencies is such that the output at each frequency is less than that which would be obtained if only one of the signals were present. In fact, the total output for *both* signals is generally less than the sum of the outputs for each separately. The practical result is that the effective gain for one fixed input is decreased when another signal is introduced, if both are in the saturation region. A typical curve showing this effect is given in Figure 26-10. Here, the input power P_1 at frequency f_1 remains constant at a value
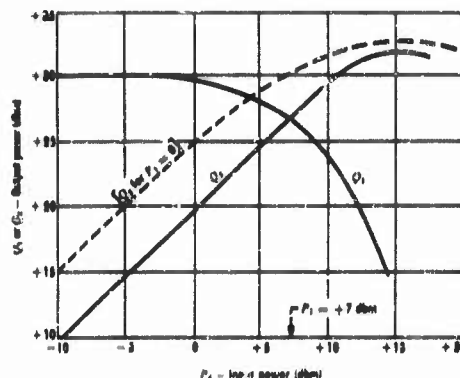
FIGURE 26-10   Two-signal response of a traveling-wave amplifier

corresponding to an output 3 db below saturation. Input power P_2 at f_2 is varied from zero up to saturation, and is plotted as the abscissa. The two output powers Q_1 and Q_2 at f_1 and f_2 are shown as the ordinate. Also shown (dotted) is the saturation curve for signal 2 alone ($P_1 = 0$). It can be seen that the signal at frequency f_1 is suppressed more and more as the other signal increases.

26.3.7 Power Tubes

Large amounts of c-w power can be obtained from traveling-wave tubes with reasonable efficiencies, over relatively large bandwidths. The primary limitation on the output power of a TWT is the dissipation ability of the circuit, which is relatively low in the case of the broadband helix. Larger output powers can be obtained, then, either by going to another kind of circuit, which invariably reduces the bandwidth, or by reducing the average dissipation by going to pulsed operation, or both.

The efficiency of a TWT, which is defined as the ratio of maximum output-signal power to beam power, depends critically upon several design parameters. Efficiencies as high as 50 percent have been achieved, with values from about 10 to 35 percent being typical for practical tubes.

Continuous-wave output powers obtainable in traveling-wave tubes range up to about 200 watts at S-band, with commercially-available tubes presently in the range of 1 to 100 watts. At X-band, tubes are at present available at the 100-watt level.

Under pulsed conditions, output powers up to a megawatt at S-band have been obtained in a commercial tube.

26.3.8 Low-Noise Receiving Tubes

The noise output of beam-type devices ordinarily originates at the heated cathode, where electrons are emitted with a random distribution of velocities. Special techniques have evolved to reduce the amount of noise which originates in traveling-wave tubes, and it has been possible to obtain very low noise figures by careful attention to the design of the gun and accelerating regions of the tube.

Although the factors which determine the lower limit to the noise figure attainable in beam-type microwave tubes are not yet completely understood, it has been found that the condition of the cathode surface is of some importance to low-noise performance. Thus, careful fabrication, construction and processing of the tube are essential, as well as proper design of the accelerating-potential profile and interaction region of the tube. In operation, the noise figure of a given tube depends on the proper adjustment of cathode temperature, electrode voltages, and the alignment of the tube in the focusing field.

At present, commercial tubes are available at S-band whose noise figures are as low as 2 to 3 decibels, with gains about 25 decibels, and outputs near one milliwatt. At X-band, the noise figure is about 5 decibels, with lower values having been measured in the laboratory.

26.3.9 Modulation and Control

Amplitude or phase modulation can be produced in a traveling-wave amplifier by appropriate variations of the beam current or beam voltage. This property makes possible a number of attractive master-oscillator power-amplifier applications, in which the desired modulation of the signal can be accomplished, without the usual difficulties which arise when an oscillator is modulated directly.

Amplitude modulation results when the beam current is varied, usually by means of a control grid in the gun, while a constant signal is applied at the input. Incidental phase modulation is also produced in this case, and thus the method may be limited to those applications where the r-f phase is relatively unimportant.

Phase modulation occurs when the helix voltage changes, because of the change in the velocity of the beam, and therefore, the number of wave-

lengths on the tube. Since the gain of the tube depends rather critically on synchronism between the wave and the beam, the helix voltage can be changed only by a small amount before incidental amplitude modulation becomes serious. In a typical traveling-wave tube, a phase deviation of about 360 degrees can be obtained, with only a few decibels change in the output level; this is for a helix-voltage change of 50 volts.

While the phase deviation may be limited to about 360 degrees, an unlimited deviation can be effectively simulated by applying a sawtooth waveform to the helix, such that the 360 degrees shift is continuously repeated. The TWT operated in this mode is called a *serrodyne*. The effect is to add or subtract one full cycle of the r-f signal during each period of the sawtooth, and thus the frequency of the output signal is shifted with respect to the input frequency, the amount of the shift being equal to the frequency of the sawtooth. Under ideal conditions, when the peak voltage change on the helix is exactly that which produces 360 degrees phase shift, and the flyback time of the sawtooth is zero, the spectral content of the output consists purely of the shifted signal. In a practical case, the shifted signal can be made more than ten decibels greater than any other component. The serrodyne finds application in doppler simulation, single-sideband generation, and wideband frequency modulation, among others.

A wide variety of combinations of amplitude and phase modulation is possible with a TWT, which allows the synthesis of practically any kind of microwave spectral distribution within the bandwidth limitations of the tube and its modulating electrodes. For example, balanced, or suppressed-carrier modulation can be accomplished by amplitude-modulating with a full-wave-rectified signal, and simultaneously reversing the phase with a square-wave version of the signal applied to the helix. See Figure 26-11.

Thus, the traveling-wave amplifier is an extremely flexible device for producing a wide range of different kinds of microwave signals, and is capable of accomplishing this over a large frequency range without need for any mechanical tuning adjustments.

In certain applications, the transient response of a TWT is of importance. Although the transient behavior is quite complex in such a distributed-interaction device, some general statements about the rise times and delays can be made. First, the delay times for forward-wave amplifiers are of the order of magnitude of the time required for the signal to travel down the slow-wave structure, which in turn is approximately the same as the beam transit time.

If the signal input is pulsed on while the beam remains on, the rise time of the output signal depends on the steady-state bandwidth characteristics of the tube: for broadband, nondispersive tubes, the output rise time is very
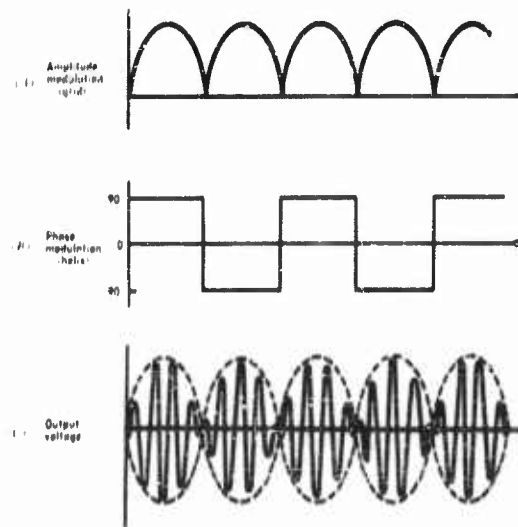
FIGURE 26-11 Synthesis of a suppressed-carrier waveform by simultaneous amplitude and phase modulation of a traveling-wave amplifier

small, of the order of one or two cycles of the r-f signal; for narrow-band dispersive tubes, the rise time is larger by perhaps an order of magnitude, depending on the details of the steady-state response and the tube parameters. In a typical dispersive tube, the rise time is about equal to the beam transit time.

When a continuous signal is applied to the input, and the beam is pulsed on, several simultaneous phenomena take place, which again depend on the bandwidth characteristics of the tube. Rise times typical of S-band tubes are from perhaps 15 to 50 millimicroseconds. A pulse 1 beam tends to excite video-frequency responses on the circuit, and if the input and output couplers are not well matched at video frequencies (as is usually the case) the circuit will "ring", producing an oscillatory distortion on the r-f output envelope. This effect will be minimized in a tube intended for pulsed-beam operation; otherwise, these transients can persist for several hundred millimicroseconds. In a typical tube intended for pulse modulation of S-band signals, the rise time is 20 mμsec and the delay about the same.

26.4 Backward-Wave Oscillators and Amplifiers

Backward-wave devices are similar in many respects to the traveling-wave amplifier discussed above. The interaction process takes place between the electron beam and a slow wave on the circuit which travels parallel to

the beam and exerts longitudinal forces on the electrons. Bunching of the beam takes place, and the kinetic energy of the beam is converted into a-c power on the circuit. However, in contrast with the traveling-wave amplifier, the backward-wave tube is inherently regenerative. The difference is due to the fact that the beam interacts with a *backward* wave, that is, a wave with oppositely-directed phase and group velocities. To obtain interaction the electron velocity must be synchronized with the *phase* velocity of a wave on the circuit, and the result is that the direction of the flow of r-f energy on the circuit is opposite to that of the electrons, resulting in a built-in feedback mechanism.

An important property of backward waves is that the phase velocity is a strongly-varying function of frequency; i.e., backward waves are inherently dispersive. A direct result is that, for a given beam velocity, interaction can take place only over a narrow band of frequencies, and, for fairly weak interaction, the device acts as a narrow-band regenerative amplifier. Since the frequency at which this interaction takes place depends on the electron velocity, the amplifier can be continuously tuned by changing the beam voltage. For stronger interaction (more beam current), the internal feedback mechanism causes the device to break into oscillation at a frequency again determined by the beam velocity. Thus, the backward-wave tube has the unusual property of being voltage-tuned. From the standpoint of countermeasures applications, this is a most important feature, since it makes possible rapid tuning of microwave oscillators and amplifiers, thus opening the door to a new class of systems based on rapid tunability over broad bandwidths.

26.4.1 Mechanism of Operation

The operation of backward-wave tubes can be understood by considering
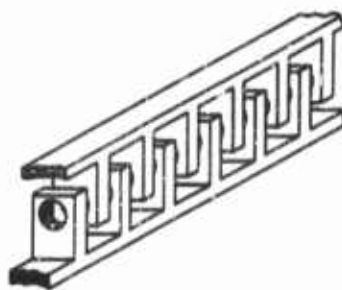
FIGURE 26-12 Folded-line circuit for a backward-wave tube

a simple structure on which a backward wave can propagate, such as the folded transmission line shown in Figure 26-12. This structure can be viewed as a strip line or a waveguide, winding back and forth upon itself, with a series of holes through which the electron beam can pass. For a TE mode on the guide, the electric field will then be directed along the line of motion of the beam at the points where the beam crosses the guide, and thus the beam "sees" longitudinal accelerating and decelerating fields, separated by drift regions in which there are no fields.

Assuming a wave exists on the structure, and is traveling from right to left, we can determine the effect of this wave on the electrons as they encounter the fields at the successive gaps. Figure 26-13a shows the situation
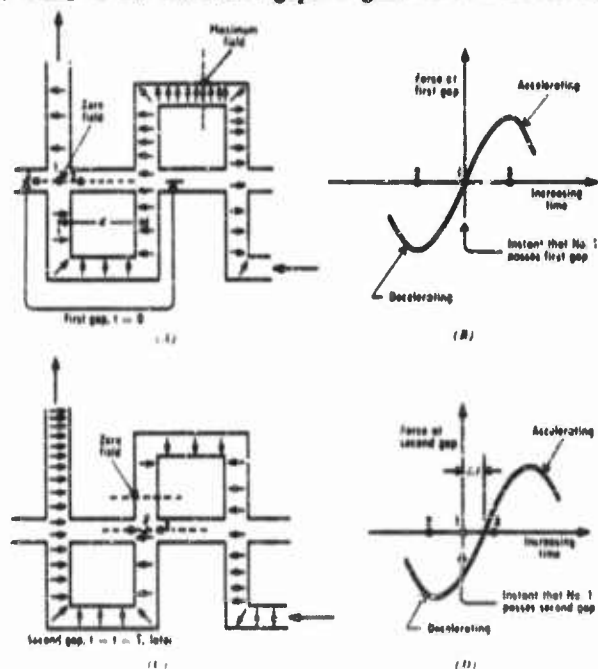
FIGURE 26-13   Diagram illustrating the interaction process in a backward-wave tube

at the first gap at the instant when the electric field is zero, and going accelerating. (The arrows indicate the direction of the force on an electron.) At this instant, a reference electron 1 passes through the gap, experiencing no force. An earlier electron 2 had encountered a decelerating field, and had been slowed down; a later electron 3 will encounter an accelerating field, and be speeded up. Thus, velocity modulation occurs at this gap, just as in a klystron gap. It can be seen that both 2 and 3 will move toward 1, and thus electron 1, our reference, will be in the center of a bunch.

Next, consider the situation when our reference electron reaches the second gap. This occurs at a time $T_s = d/u_0$ seconds later, where u_0 is the electron velocity, and d is its path length between gaps. If, during this same time, the wave on the structure travels with such a speed that the *next* zero just arrives at the second gap, then the reference electron will see exactly the same field at gap 2 as it saw at gap 1, i.e., zero field, going accelerating.

In order to do this, the wave must travel just L meters short of a half wavelength, or $\lambda_0/2 - L$ meters. If the wave velocity on the line is c, the time required to travel this distance is $\lambda_0/2c - L/c$ seconds. Actually, the velocity is adjusted so that the reference electron arrives at the second gap slightly before the zero of field arrives, so that it encounters a slight deceleration. This situation is illustrated in Figure 26-13b. The time, $\triangle t$, by which the electron "beats" the zero of field to the second gap is

$$\triangle t = [\lambda_0/2c - L/c] - T_e = 1/2f - L/c - d/u_0$$

As this situation is repeated from gap to gap, the bunch around the reference electron continues to become more compact; it also slows down, giving up energy to the circuit wave, as long as it continues to see a decelerating field at each gap. This last condition can be satisfied by adjusting the velocity such that, after passing through n gaps, the bunch is still within a half cycle of the zero of field. That is, $n\triangle t$ is approximately one-half of the period of the r-f wave. Under these conditions, the electron bunches encounter strong decelerating fields over the entire length of the tube, and a maximum of the beam energy is converted into wave energy on the circuit. The frequency at which this strong interaction takes place can be found by setting $n\triangle t$ equal to a half period of the wave, or:

$$1/2f = n\triangle t = n/2f - nL/c - nd/u_0.$$

Solving for the frequency f,

$$f = \frac{[1 - 1/n]}{2[L/c + d/u_0]}.$$

This is the basic relation for the frequency of a backward-wave tube using the interdigital-line circuit, in terms of the dimensions and the beam velocity.

As the beam loses energy to the wave, the wave tends to grow in amplitude. But, since the backward wave, with which the beam is interacting, carries power in a direction opposite to the motion of the beam, the effect is for the wave to grow towards the gun end of the tube. This causes even stronger bunching, and the result is a built-in feedback of a regenerative nature. For a given physical length of tube, there is a critical beam current, called the "starting current", above which the regeneration is strong enough to cause the tube to oscillate at about the frequency given in the above expression. Below starting current, a signal which is introduced at the input (collector end of the tube) will be amplified if its frequency is near this value. As with all regenerative amplifiers, the gain can be very large, but is critically dependent on beam current and signal frequency.

26.4.2 Slow-Wave Structures

Any structure which will support a backward wave can potentially be used for backward-wave interaction. The folded line is one of the simplest in concept, and is capable of high power dissipation. It can take the form of a folded strip line, as shown in Figure 26-12, or the sides can be closed off, in which case the line is a folded waveguide, operating in the TE_{10} mode. Intermediate cases, in which the enclosure is further away, can be recognized as folded ridged (or loaded) waveguide.

It can be shown that any waveguide with periodic variations in the direction of propagation can support backward waves in the form of space-harmonic components of the total field. Thus, a wide variety of circuits have found application in backward-wave devices, and, unfortunately, many circuits for FWAs seem to prefer backward-wave oscillations under conditions of high-power operation.

The helix is an excellent circuit for backward-wave tubes in the frequency range where the circumference is one-half to one-fourth the free-space wavelength. Ordinarily, such a helix is wound of tape, with the tape width about equal to the gap between turns. Under these conditions, the fields seen by a beam quite close to the helix are similar to those of the folded-line structure, with regions of axial field between turns alternating with drift regions beneath the tape. It is customary to use a hollow beam in a helix-type backward-wave tube, since only those electrons near the helix can participate in the interaction. It should be noted that, for a given frequency of operation, the diameter of the helix for a backward-wave tube is several times larger than for a FWA. Thus, for helix-type tubes, backward-wave interaction can be utilized at higher frequencies than can forward-wave interaction.

As with forward-wave tubes, it is important that reflections at the ends of the circuit be minimized to reduce undesirable regenerative effects. In backward-wave tubes, the effect of imperfect matches is to produce sinusoidal variations in the starting current and power output as frequency is varied. A related effect is frequency pulling, or a shift in the oscillator frequency with changing VSWR of the load. If the matches at both ends of the tube are better than about 2:1 VSWR, and some additional loss is introduced on the circuit, these effects are ordinarily tolerable. The pulling figure of a BWO is generally lower than that of conventional oscillators. Loss has detrimental effect on the efficiency, and hence should be minimized.

Another advantage of the helix circuit over the interdigital line is that it is somewhat easier to obtain a good broadband match to the helix.

26.4.3 Guns and Focusing

Generally the requirements for production and maintenance of beams for

O-type backward-wave tubes are the same as for forward-wave tubes. The most significant difference is that backward-wave tubes operate over a wide range of voltages, rather than a fixed voltage, and must be so designed that the required current for oscillation is supplied over the entire range of operating voltages. The variation of voltage and current as a function of frequency is such that the required beam perveance is normally highest at the lowest frequency of operation, and this point determines the gun design. At higher frequencies, the beam is simply accelerated by a voltage difference between the circuit and the gun anode.

Since the structure in a backward-wave tube is usually shorter than that of a forward-wave tube, the focusing problem is simpler, for a given frequency and power level.

Because of the variable-voltage requirement, the focusing of backward-wave tubes is almost universally of the confined-flow or Brillouin-flow types. In interdigital-line tubes with solid beams, the use of Brillouin focusing allows convergent-geometry guns and thus higher current densities and powers are attainable. Periodic electrostatic focusing is possible with the interdigital circuit by operating the two halves at different d-c potentials, but this method is restricted to very low power levels.

26.4.4 Tuning Characteristics

The frequency at which amplification or oscillation will occur in a backward-wave tube is primarily determined by the dispersion characteristic of the circuit and the beam velocity. For a given beam voltage, interaction can occur only at the frequency for which the circuit phase velocity is in approximate synchronism with the beam, and thus the "tuning curve", or a plot of frequency versus beam voltage, can be obtained directly from the dispersion curve of the circuit. A typical dispersion curve for periodic slow-wave circuit is shown in Figure 26-14, in the form of a plot of frequency ω versus the phase constant β. (On this ω-β diagram, the phase velocity of the circuit wave is given by the slope of the line from the origin to a given point on the curve, and the group velocity is the incremental slope of the curve at the given point, both velocities relative to the velocity of light.) Over the range of the curve from A to B, the phase velocity is positive while the group velocity is negative, which is characteristic of a backward wave; this is the region of operation of backward-wave devices. A plot of the relative phase velocity versus frequency is shown in Figure 26-15. If the beam at voltage V is to be synchronous with the wave, the voltage must satisfy the relation $V = 505 \, (v_p/c)^2$, and from this the synchronous tuning curve of Figure 26-16 is obtained.

In practice, the synchronous tuning curve gives only an approximate indi-
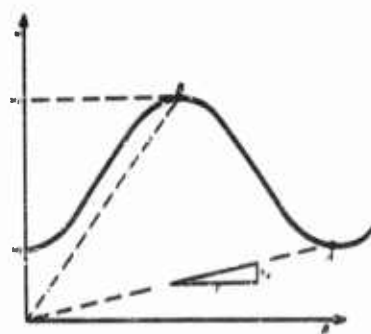
FIGURE 26-14 Frequency-vs-phase constant curve for a periodic slow-wave structure
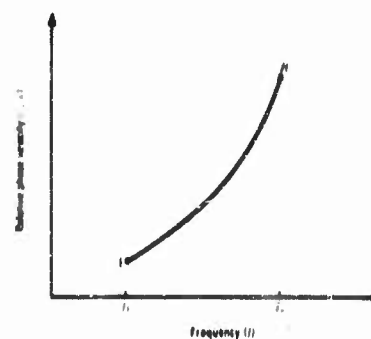
FIGURE 26-15 Phase velocity-vs-frequency for a backward wave
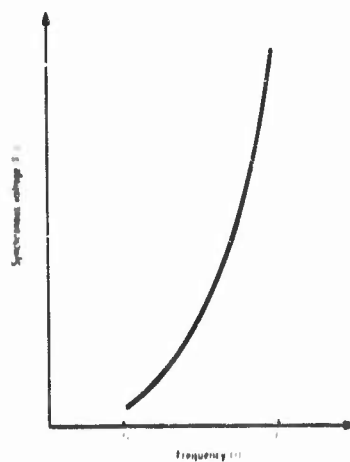
FIGURE 26-16 Synchronous tuning curve for a backward-wave tube
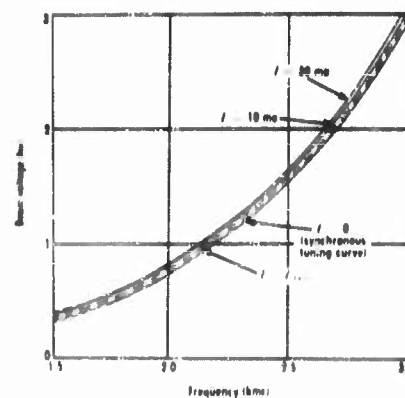
FIGURE 26-17 Typical tuning characteristics of a backward-wave oscillator, illustrating frequency pushing due to changing beam current

cation of the performance, and must be corrected by including the effects of other factors to obtain the actual tuning characteristic. The most important of these factors is the beam current, and its effect on the frequency of oscillation is termed "frequency pushing." As the beam current increases, the frequency decreases, for a fixed voltage. This effect is illustrated on the

typical tuning characteristics shown in Figure 26-17, in which several curves are plotted, for various values of beam current.

The amount of frequency pushing which will occur in a given tube depends on beam current and voltage, and thus varies over the range of tube operation; values up to about five percent shift from the start-oscillation frequency are typical for high-power tubes.

Frequency pulling, mentioned in Section 26.4.2, results in minor variations in the tuning curve, c: y variations with frequency in the reflections at the ends of the circu... _he amount of pulling also depends on the match at the load, and the length of the connecting transmission line. For relatively good matches, the deviations from the smooth tuning curve are approximately sinusoidal, with peak values usually less than 1 percent shift in frequency. More serious pulling and discontinuous jumps in the tuning curve can result if the termination at the collector end of the circuit is particularly bad. Such a breaking up of the tuning curve is a result of "long-line effects", which cause rapid variation in the phase of the reflected wave between the load and the collector end of the circuit.

The tuning characteristic of an O-type BWO is inherently nonlinear; typically, a 2 to 1 range in frequency can be tuned by varying the beam voltage over a 10 to 1 or smaller range.

26.4.5 Power and Efficiency

Power outputs of the order of milliwatts can readily be obtained from BWOs at frequencies as high as 100,000 megacycles per second. Larger powers can be achieved at lower frequencies, up to 10 kilowatts at S-band.

Efficiencies attainable in BWOs are generally somewhat lower than those of TWTs, being of the order of 20 percent and higher for high-power tubes, and a few percent or less for low-power tubes.

26.4.6 Modulation and Control

Possibilities for modulating a BWO include variation of the beam voltage, or the beam current; in addition, there is the possibility of injecting a signal at the collector end of the circuit.

The primary effect of varying the beam voltage is to change the frequency of the output signal. Thus, the BWO can be frequency-modulated by applying an appropriate signal waveform on the circuit. Due to the nonlinear shape of the tuning curve, the modulating signal must be predistorted if a linear modulation characteristic is desired. The limit on the tuning rate that can be achieved is set by the build-up time of the oscillations, which is the order of several transit times of the beam. Since the transit time is usually measured in millimicroseconds, tuning rates up to about 1000 megacycles per second per microsecond can be obtained in a typical S-band tube.

Since the output power, as well as the frequency, changes with beam voltage, some amplitude modulation is generally present in the output, when the beam voltage is modulated. If the frequency deviation is small, however, the AM can be negligible.

If the beam voltage is fixed and the current varied, the primary effect is to vary the power output. Thus, a BWO can be amplitude-modulated by varying the beam current. Because of the frequency-pushing effects discussed previously, some incidental FM is also produced when the current is modulated. Usually a special gun design is required to vary the beam current, e.g., with a control grid.

In relatively low-power applications, it is usually more convenient to produce AM by some external means, as for example, with a diode or ferrite modulator. In addition to reducing the effects of pushing, such a device can have a relatively constant modulation sensitivity over the frequency range of operation, which would be difficult to obtain by direct modulation of the BWO beam current.

The output signal of a BWO can be frequency-controlled by the injection of a signal into the (normally terminated) collector end of the circuit. If the injected signal is near the frequency of oscillation, the oscillation can become locked to the signal frequency. The behavior of a BWO under these conditions is similar to that of other locked oscillators, and depends on the strength of both the oscillation and the injected signal, as well as the difference in their frequencies. If the BWO is operating at just slightly above starting current, an injected signal near the oscillating frequency will be amplified, while the oscillation will be simply suppressed, without any frequency pulling. The resulting signal, locked on the frequency of the injected signal, is larger than would be obtained from the oscillator by itself.

At currents considerably above the start-oscillation value, the behavior is somewhat more complicated, with extra-frequency components being generated, and pulling of the oscillator frequency occurring. The range over which locking takes place decreases as the current is raised above the starting value, and typically might be a few megacycles wide.

The locking-in behavior of a typical low-power BWO is illustrated in Figure 26-18, which shows the frequency spectrum of the output. In (a), the BWO is barely oscillating at f_0, and a constant-amplitude injected signal at f_s is tuned toward f_0. As the two frequencies get closer, the injected signal simply grows, while the oscillation is suppressed. In (b), the current is about 1.5 times starting current. Here, the oscillator frequency is pulled toward that of the injected signal, and equally-spaced extra-frequency components appear on one side of the oscillator frequency in the pulling region. Finally, lock-in does occur, but over a narrower range than in (a).

(A)   (B)
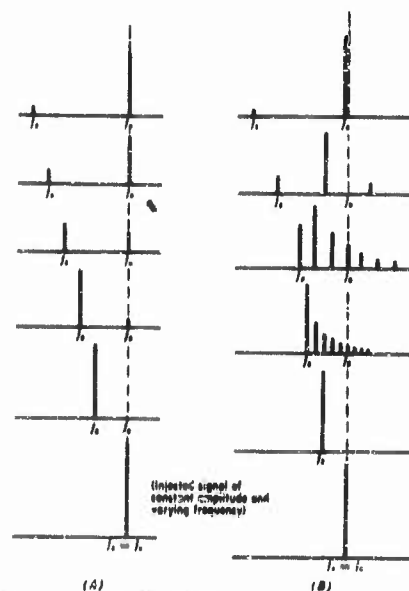
FIGURE 26-18 Spectrum of a backward-wave oscillator in the presence of an injected signal

26.4.7 Spurious Outputs

Under certain conditions of operation, a BWO can produce outputs at frequencies other than that of the fundamental or main output. These extra outputs are termed spurious outputs, and are generally detrimental to the performance of a system in which the BWO is used.

It was seen in the last section that, in the presence of a locking signal, spurious outputs can occur when the oscillator is on the verge of locking in, but these disappear under the locked condition.

Another possible source of spurious outputs in an O-type BWO is in modes corresponding to higher-order solutions of the start-oscillation condition. Generally, the tube can oscillate on these modes if the beam current is sufficiently high, of the order of five times the fundamental starting current or higher. When this condition is satisfied, the higher-mode output appears at a frequency somewhat lower than that of the fundamental output, and the strength of the spurious output is usually much smaller than the fundamental output. Because of the nonlinear character of the beam, intermodulation products appear whenever two signals are present, so that the result is that a multiple distribution of spurious signals is observed when the BWO is oscillating simultaneously on two modes. Higher-mode operation can be eliminated simply by reducing the physical length of the backward-wave circuit, and therefore should not be a problem in well-designed tubes.

Depending on the kind of circuit employed, interaction may take place with various space-harmonic components of the field other than the fundamental backward wave, to produce spurious outputs. This kind of interaction can be virtually eliminated by proper circuit design.

Finally, spurious outputs occur at the harmonics of the fundamental frequency due to the nonsinusoidal bunching of the beam. These outputs are usually outside the frequency range of interest, as well as being relatively small in amplitude.

In low-power applications, such as local oscillators and signal generators,

In which spurious outputs are a most serious problem, the O-type BWO is capable of extremely good performance, producing a clean spectrum in which the spurious signals are typically 70 to 80 decibels or more below the main output.

26.4.8 Low-Noise Backward-Wave Amplifiers

Since the backward-wave amplifier is an electronically-tunable, narrow-band amplifier, its most important application would seem to be as the input stage for a rapidly-tuned search receiver. In such an application, the sensitivity attainable depends on the minimum noise figure of the backward-wave tube. It has been recently demonstrated that these tubes can be designed and constructed with noise figures as low as 2.5 decibels, with gain well over 30 decibels, which appears to establish the BWA as an important component in microwave systems.

26.5 Power and Size Limitations

In assessing the potentialities of microwave tubes for specific new applications, the problem is generally one of determining what can be accomplished, within the state-of-the-art, toward meeting the requirements of power, bandwidth, and physical size at a specific operating frequency. Basic to an understanding of how the performance characteristics depend on frequency are the laws of *scaling*; this section will be devoted to a statement of general scaling rules, and their implications with regard to the power and size limitations of TWTs.

26.5.1 Scaling Laws

Suppose that we have a successful design of a microwave tube, and wish to investigate the design of a similar tube to operate at a different frequency or power level. We should expect to have to change the physical dimensions, voltages, currents, and the magnetic field, as well as the operating frequency. The problem is to determine the new values of the performance characteristics, in terms of these changes. In general, two requirements must be met in the scaled version, and these requirements yield the scaling laws which must be satisfied: (1) the shapes of the electron trajectories must remain unchanged with scaling; and (2) the electrons must traverse corresponding distances in the same fraction of a cycle of the operating frequency.

Consider two geometrically-similar microwave tubes whose linear dimensions are in the ratio L:

$$L = L_2/L_1$$

These tubes are to operate at frequencies in the ratio F:

$$F = f_2/f_1$$

If we now consider the dynamics of the electron beam in the electric and magnetic fields present, and impose the conditions that trajectory shapes and transit angles remain unchanged with scaling, we can derive from the equations of motion the general scaling ratios between corresponding quantities in the two tubes. The results are summarized in Table 26-I, under the column

TABLE 26-I. SCALING LAWS FOR MICROWAVE TUBES

Based on: (1) Geometrically similar electron trajectories
(2) Identical transit angles

Scaling-ratio definitions: (1) Length: $L = L_2/L_1$
(2) Frequency: $F = f_2/f_1$

Quantity	General Scaling	Complete Scaling $LF = 1$	Dimension Scaling $F = 1$
Wavelength, λ_2/λ_1	$1/F$	$1/F = L$	1
Time, dt_2/dt_1	$1/F$	$1/F = L$	1
Distance, ds_2/ds_1	L	$1/F = L$	L
Area, A_2/A_1	L^2	$1/F^2 = L^2$	L^2
Velocity, v_2/v_1	FL	1	L
Acceleration, a_2/a_1	F^2L	$F = 1/L$	L
Electric field, E_2/E_1	F^2L	$F = 1/L$	L
Magnetic field, B_2/B_1	F	$F = 1/L$	1
Potential, V_2/V_1	F^2L^2	1	L^2
Charge density, ρ_2/ρ_1	F^2	$F^2 = 1/L^2$	1
Current density, i_2/i_1	F^3L	$F^3 = 1/L^2$	L
Current, I_2/I_1	F^3L^3	1	L^3
Power density, p_2/p_1	F^5L^3	$F^2 = 1/L^2$	L^3
Power, P_2/P_1	F^5L^5	1	L^5
Perveance, S_2/S_1	1	1	1

headed "general scaling." This column applies to tubes with different dimensions *and* operated at different frequencies, and shows how voltage, power, current density, etc., in the *scaled* tube compare with those quantities in the *original* tube. In addition, the general scaling ratios allow one to derive a number of scaling invariants, or combinations of parameters which do not change with scaling. For example, $(I_2/I_1)/(V_2/V_1)^{3/2}$ is such an invariant, since this particular ratio is independent of both L and F. It will be recognized that this is just the ratio of perveances, S_2/S_1, and thus the perveance is not a function of the scale. Another example might be the ratio V/B^2L^2.

The remaining two columns of Table 26-I are for special cases of scaling

procedure. "Complete" scaling refers to the case in which the dimensions are scaled in direct proportion to the operating wavelength, or in inverse proportion to the operating frequency. In this case, $L = 1/F$, and the ratios in this column show how the various parameters change with either L or F. For example, if the dimensions of a tube are all reduced by a factor of four, in order to operate at four times the frequency, the voltages, current, and powers remain unchanged. However, the electric and magnetic fields required are four times as large in the smaller tube, and the charge density, current density, and power density are sixteen times larger. Since the heat-dissipating ability of a structure is related to the power density the smaller tube is likely to be severely limited because of the high power-density ratio. Further, the current density obtainable from the cathode of the electron gun is limited, and thus the extent to which complete scaling can be carried is restricted by these two considerations. "Complete" scaling is the case of most practical usefulness in microwave-tube considerations.

If the frequency of operation remains fixed, and the dimensions are changed, say, to obtain more power, then the ratios in the column headed "dimension scaling" apply. If the size of the tube is doubled, and appropriate changes made in the other parameters, the power will be increased by 32 times. In this case, the magnetic field remains the same; the electric fields and current density are doubled; the voltages must be increased by a factor of four; and the power density and current are increased eightfold. Scaling on this basis will again be limited by allowable power dissipation, available current density, or voltage breakdown.

It is important to note that, while the above scaling relations were derived from the equations of motion of the electron beam alone, the results are directly applicable to such a complicated device as a TWT, if *complete scaling* is employed. The reason for this is that the performance of such a tube depends on the choice of a number of design parameters which are *inherently independent* of this kind of scaling. For example, in a helix-type TWT, the parameter most basic to the over-all performance is the frequency-size parameter, γa, which is directly proportional to frequency and helix radius, and inversely proportional to the square-root of the beam voltage. Examination of the complete scaling column shows that γc is an invariant in this kind of scaling. Other design parameters, which depend on γa, and are also invariant with complete scaling are: C, QC, N, B, BCN, etc. It follows that the gain, bandwidth, and efficiency are also invariant. Thus, any conclusions drawn about how the *beam* power varies with scaling are also applicable to the r-f power output of the tube, at least to a first approximation. Some reservation is warranted here, though, since the efficiency depends also on some factors which may not be entirely independent of scal-

ing, such as the loss on the helix. Also, it should be noted that the scaling relations apply to the values of the "design center" parameters, and should not be used, for example, to calculate how the power output varies across the band of operation of a given traveling-wave tube.

26.5.2 Size Limits at Low Frequencies

If one attempts, by complete scaling, to scale an S-band TWT down to about fifty megacycles per second, the result would be a tube perhaps forty feet long and a foot in diameter; such a tube would be less than useful. Even at 500 megacycles per second, a four-foot tube would be impractical in many applications. Furthermore, in this frequency region, conventional-tube devices are competitive with traveling-wave tubes, the prime example being the distributed amplifier. Thus, there is a frequency limit below which the TWT becomes too large to be practical.

The low-frequency limit of practical traveling-wave tubes depends on what is considered a tolerable physical size, in any particular application, but probably lies in the range from 50 to 500 mcs. The limit can be lowered for a given size by designing the tube to utilize a hollow beam, thus increasing the attainable gain per wavelength, which shortens the required physical length at a given frequency.

A developmental traveling-wave amplifier has been reported which operates over the range of 50 to 300 mcs, with power output at the 1000-watt level, whose over-all physical length is about thirty inches. It would seem to be difficult to do much better than this with TW tubes.

26.5.3 Power Limits at High Frequencies

The performance of a traveling-wave device at high frequencies is limited primarily by three factors: the dissipation ability of the circuit, current-density capability of the gun and focusing system, and the difficulty of fabricating small tube parts. If complete scaling of a dissipation-limited helix-type TWT is attempted, the attainable output power will drop off as the square of the wavelength. However, ordinary construction techniques begin to fail at frequencies above about 30,000 mcs, and it has proved difficult to obtain even the scaled powers at these frequencies. With special techniques, the high-frequency limit for helix-type amplifiers may be about 50,000 mcs, and perhaps as high as 100,000 mcs for BWOs. Achievable powers of tens of milli-watts seem to be the limit in this frequency range. Several alternative structures for BWOs have been developed, which extend the frequency limits upward by perhaps a factor of 2.

In contrast to the situation at low frequencies, there are no alternative devices which begin to compete as the high-frequency limits of traveling-wave devices are approached.

26.5.4 Limits Imposed by Gun and Focusing System

The current densities attainable in an electron beam are dependent on the capabilities of the gun and focusing system. The cathode surface itself is subject to fairly definite limits on emission-current density, consistent with reasonable tube life. Continuous d-c emission densities of about 0.2 amp per cm squared can reliably be obtained from ordinary oxide cathodes, and up to about 5 amp per cm squared from the so-called "dispenser" or impregnated cathode. For short pulses, less than about five microseconds, current densities up to about 10 amp per cm squared can be drawn from either type of cathode without serious reduction of tube life.

These limits on primary current density are considered severe by the designers of high-power microwave tubes; higher values would be desirable in many applications. Effective beam-current densities up to the order of ten times the above values can be obtained by converging the beam in the gun region, so that a small beam is produced from a large cathode. Here, the limit to the amount of convergence which can be attained is set by random "thermal" velocities in the transverse direction, which tend to smear the beam out sideways, preventing the attainment of large convergence ratios.

Another factor in gun design which limits the amount of current that can be obtained at a given voltage is the perveance, or current per $(volt)^{3/2}$. The perveance of a beam is limited because of the effects of space-charge forces; in Pierce-type guns, perveances of about 2×10^{-6} amp per $(volt)^{3/2}$ seem to be the upper limit, for solid cylindrical beams.

In high-perveance, high-convergence electron guns, another related effect occurs which tends to limit the performance. The size of the anode aperture becomes comparable to the cathode-anode spacing, and the fields in the gun suffer considerable distortion. The results are: nonuniform emission over the cathode surface, larger beam size, lower total current, and appreciable spherical aberration.

Once the desired beam has been formed by the gun, the focusing system must be capable of confining the flow to the specified region by counteracting the normal space-charge spreading forces. For high-power, high-perveance beams, confined-flow or Brillouin-flow focusing is generally required, although any of the methods discussed in Section 26.3.4 are applicable within their limitations. For example, periodic-magnetic focusing is, at the moment, limited by the amount of coercive force of available magnetic materials to tubes delivering about 1 to 10 watts of output power.

Finally, it should be noted that the focusing system must be compatible with the gun design, so that the beam flows smoothly into the circuit region of the tube. If a strong confining magnetic field is used for focusing the beam, a converging geometry cannot be used in the gun, unless the focusing

field is likewise convergent in the gun region, or else eliminated entirely by a magnetic shield.

26.6 Bibliography

26.6.1 General Material on Traveling-Wave Tubes

D. A. Dunn, "Traveling-wave amplifiers and backward-wave oscillators for VHF," IRE Trans., (Electron Devices) vol. ED-4, no. 3, July, 1957, pp. 246-264.

R. G. E. Hutter, "Traveling-wave tubes," Advances in Electronics and Electron Physics, vol. 6, 1954, pp. 371-401.

J. R. Pierce, "Traveling-wave tubes," D. Van Nostrand Co., Inc., New York, N. Y., 1950.

J. R. Pierce, "Some recent advances in microwave tubes," Proc. IRE, vol. 42, no. 12, December, 1954, pp. 1735-1747.

F. E. Terman, "Electronic and radio engineering," McGraw-Hill Book Co., Inc., New York, 1955, pp. 678-688.

26.6.2 Backward-Wave Tubes

W. V. Christensen and D. A. Watkins, "Helix millimeter-wave tube," Proc. IRE, vol. 43, no. 1, January, 1955, pp. 93-96.

M. R. Currie and D. C. Forster, "The gain and bandwidth characteristics of backward-wave amplifiers," IRE Trans., (Electron Devices) vol. ED-4, no. 1, January, 1957, pp. 24-34.

M. R. Currie and J. R. Whinnery, "The cascade backward-wave amplifier: a high-gain voltage-tuned filter for microwaves," Proc. IRE, vol. 43, no. 11, November, 1955, pp. 1617-1631.

R. W. Grow and D. A. Watkins, "Backward-wave oscillator efficiency," Proc. IRE, vol. 43, no. 7, July, 1955, pp. 848-856.

H. R. Johnson and R. D. Weglein, "Backward-wave oscillators for the 8000-18,000 megacycle band," IRE Trans., (Electron Devices) vol. ED-4, no. 2, April, 1957, pp. 180-184.

H. R. Johnson, "Backward-wave oscillators," Proc. IRE, vol. 43, no. 5, May, 1955, pp. 684-697.

A. Karp, "Traveling-wave tube experiments at millimeter wavelengths with a new, easily-built, space harmonic circuit," Proc. IRE, vol. 43, no. 1, January, 1955, pp. 41-46.

A. Karp, "Backward-wave oscillator experiments at 100 to 200 kilomegacycles," Proc. IRE, vol. 45, no. 4, Ap. l, 1957, pp. 496-503.

P. Palluel and A. K. Goldberger, "The O-type carcinotron tube," Proc. IRE, vol. 44, no. 3, March, 1956, pp. 333-345.

26.6.3 Guns and Focusing

R. Adler, O. M. Kromhout, and P. A. Clavier, "Transverse-field traveling-wave tubes with periodic electrostatic focusing," Proc. IRE, vol. 44, no. 1, January, 1956, pp. 82-89.

Kern K. N. Chang, "Beam focusing by periodic and complementary fields," Proc. IRE, vol. 43, no. 1, January, 1955, pp. 62-71.

K. K. N. Chang, "Confined electron flow in periodic electrostatic fields of very short periods," Proc. IRE, vol. 45, no. 1, January, 1957, pp. 66-73.

K. K. N. Chang, "Biperiodic electrostatic focusing for high-density electron beams," Proc. IRE, vol. 45, no. 11, November, 1957, pp. 1522-1527.

C. B. Crumly, "A UHF traveling-wave amplifier tube employing an electrostatically focused hollow beam," IRE Trans., (Electron Devices) vol. ED-3, no. 1, January, 1956, pp. 62-65.

C. C. Cutler and M. E. Hines, "Thermal velocity effects in electron guns," Proc. IRE, vol. 43, no. 3, March, 1955, pp. 307-315.

M. S. Glass, "Straight-field permanent magnets of minimum weight for TWT focusing-design and graphic aids in design," Proc. IRE, vol. 45, no. 8, August, 1957, pp. 1100-1105.

L. E. S. Mathias and P. G. R. King, "On the performance of high perveance electron guns," IRE Trans., (Electron Devices) vol. ED-4, no. 3, July, 1957, pp. 280-286.

J. T. Mendel, "Magnetic focusing of electron beams," Proc. IRE, vol. 43, no. 3, March, 1955, pp. 327-331.

J. R. Pierce, "Theory and design of electron beams," D. Van Nostrand Co., Inc., 2nd Ed., New York, N. Y., 1954.

26.6.4 Nonlinear Behavior of TWTs

D. J. Bates and E. L. Ginston, "A traveling-wave frequency multiplier," Proc. IRE, vol. 45, no. 7, July, 1957, pp. 936-944.

R. W. DeGrasse and G. Wade, "Microwave mixing and frequency dividing," Proc. IRE, vol. 45, no. 7, July, 1957, pp. 1013-1015.

F. B. Fank and G. Wade, "Traveling-wave tube limiters," IRE Trans., (Electron Devices) vol. ED-4, no. 2, April, 1957, pp. 148-152.

H. Kurokawa, I. Someya, and M. Morita, "New microwave repeater system using a single traveling-wave tube as both amplifier and local oscillator," Proc. IRE, vol. 45, no. 12, December, 1957, pp. 1604-1611.

26.6.5 Medium and High-Power Tubes

J. J. Caldwell, Jr. and O. L. Hoch, "Large signal behavior of high power traveling-wave amplifier," IRE Trans., (Electron Devices) vol. ED-3, no. 1, January, 1956, pp. 6-17.

M. Chodorow and E. J. Nalos, "The design of high-power traveling-wave tubes," Proc. IRE, vol. 44, no. 5, May, 1956, pp. 649-659.

M. Chodorow and R. A. Craig, "Some new circuits for high-power traveling-wave tubes," Proc. IRE, vol. 45, no. 8, August, 1957, pp. 1106-1118.

C. C. Cutler, "The nature of power saturation in traveling-wave tubes," BSTJ, vol. 35, no. 4, July, 1956, pp. 841-876.

L. W. Holmbos and M. Ettenberg, "Development of a medium power L-band traveling-wave amplifier," IRE Trans., (Electron Devices) vol. ED-4, no. 1, January, 1957, pp. 78-81.

J. P. Laico, H. L. McDowell, and C. R. Moster, "A medium power traveling-wave tube for 6,000-mc radio relay," BSTJ, vol. 35, no. 6, November, 1956, pp. 1285-1346.

O. T. Purl, J. R. Anderson, and G. R. Brewer, "A high-power periodically focused traveling-wave tube," Proc. IRE, vol. 46, no. 2, February, 1958, pp. 441-448.

J. E. Rowe, "Design information on large-signal traveling-wave amplifier," Proc. IRE, vol. 44, no. 2, February, 1956, pp. 200-210.

W. W. Siekanowicz and F. Sterze, "A developmental wide-band, 100-watt, 30 db, S-band traveling-wave amplifier utilizing periodic permanent magnets," Proc. IRE, vol. 44, no. 1, January, 1956, pp. 55-61.

P. K. Tien, L. R. Walker, and V. M. Wolontis, "A large signal theory of traveling-wave amplifier," Proc. IRE, vol. 43, no. 3, March, 1955, pp. 260-277.

26.6.6 Low-Noise Tubes

M. R. Currie and D. C. Forster, "Experiments on noise reduction in backward-wave amplifiers," Proc. IRE, vol. 45, no. 5, May, 1957, p. 690.

M. R. Currie and D. C. Forster, "Low noise tunable preamplifiers for microwave receivers," Proc. IRE, vol. 46, no. 3, March, 1958, pp. 570-579.

T. E. Everhart, "Concerning the noise figure of a backward-wave amplifier," Proc. IRE, vol. 43, no. 4, April, 1955, pp. 444-449.

H. A. Haus and F. N. H. Robinson, "A minimum noise figure of microwave beam amplifiers," Proc. IRE, vol. 43, no. 8, August, 1955, pp. 981-991.

E. W. Kinaman and M. Magid, "Very low-noise traveling-wave amplifier," Proc. IRE, vol. 46, no. 5, May, 1958, pp. 561-567.

27

Crossed Field Microwave Tubes

E. C. DENCH, J. F. HULL, J. N. OSEPCHUK, D. A. WILBUR

During World War II and for several years thereafter, an imposing amount of work was done on the development and analysis of the traveling-wave magnetron oscillator (References 1, 2, and 3). Shortly after the beginning of the war the magnetron, with mechanical tunability, became the work horse of microwave electronic warfare equipment. However, because wide and rapid tuning requirements began to develop for both radar and counter-measures, wide electronic tunability of high-power microwave oscillators began to be the dream of the microwave engineer. The development of the voltage-tunable magnetron by Wilbur et al. (Reference 4) of General Electric Company in 1949 and subsequent work on it at The University of Michigan (References 5, 6, and 7) opened up a new field for the magnetron, and it also accelerated the trend toward voltage-tunable oscillators. Meanwhile, the demonstration of the principle of traveling-wave interaction in linear-beam tubes by Kompfner (Reference 8) and Pierce and Field (Reference 9) had stimulated work on crossed-field traveling-wave devices (Reference 10). The principal of backward-wave interaction was conceived independently by Warnecke et al. (Reference 11) and Kompfner. The application of the traveling-wave principle to crossed-field interaction by use of a backward-wave structure resulted in the first demonstration of the voltage-tunable BWO (backward-wave oscillator) tube late in 1950, by Warnecke et al.; this was called the Carcinotron (Reference 11). Soon after, prototypes were developed (References 12 and 13) in the United States.

Because of the obvious advantage of voltage-*versus*-mechanical tunability, its high-power and high-efficiency capability, its versatility, and its economy of size and weight, the M-type BWO has become one of the foremost microwave ECM transmitter tubes. In European countries it has developed a commanding lead in priority of development over the conventional mechanically tuned CW magnetron. In the United States, the hydraulically tuned CW magnetron is still the chief ECM microwave production tube, but efforts to develop new equipment are centered on M-type BWO's and other new microwave devices. M-type BWO's have been dev' ed mainly for high-power CW applications where wide and rapid volt   ability is necessary. The typical ECM uses are for spot and barrage noise jammers using both frequency and amplitude modulation. Continuous-wave power output capabilities at the present development status from 100 to 40,000 mc range from 5000 watts to 50 watts, respectively, with tuning ranges of about 1.4:1. There have been pulsed M-type BWO's developed in the power range of hundreds of kilowatts, which may find use in pulse-frequency diversity radar.

An outgrowth of the M-type BWO development is the Bitermitron, a backward-wave amplifier, used mostly as a voltage-tunable locked oscillator. An M-type BWO driving a Bitermitron gives more than twice the power output than is obtainable from a single tube. The application of this combination is the same as that for the M-type BWO.

Voltage-tunable magnetrons have developed along the lines of low- and medium-power CW tubes with tuning ranges of 2·1 and power outputs up to several tens of watts from L- to X-band. Their chief ECM use is for low-power barrage and spot jammers, and for drivers in amplifier-tube chains for high-power ECM and radar equipments.

Research effort has recently been started on crossed-field forward-wave amplifiers for ECM applications. These tubes are essentially broadband traveling-wave tubes using magnetron-type interaction. Initial results of this effort are so encouraging that this device promises to become the major ECM and radar transmitter tube of the future. Power outputs greater than 750 watts cw at X-band over 15 percent bandwidth at 10-db gain and 20 to 30 percent efficiency have been obtained on initial sample tubes. The quantitative theory of operation is so well known and proved experimentally that it can safely be predicted that in the future a minimum of 1 kw cw will be obtainable at X-band over a 30 percent bandwidth at 35 percent minimum efficiency and 20-db gain. It is also likely that the bandwidth and efficiency can eventually be increased to 50 percent or more at X-band. Correspondingly higher powers, efficiencies, and bandwidths will be obtainable at lower frequencies. Since this tube is a broadband amplifier, there is no variation of the applied voltages as the frequency is changed. Therefore the jamming

tricks, whether of the noise or deception type, can be generated at low power and then amplified. Also, existing jammers can be upgraded in power by amplifying their output with the crossed-field forward-wave amplifier. The results of the research effort on this tube are too preliminary to be included in this chapter because they might be misleading as to the eventual capability of the device, but a brief theory of operation is given. A more complete discussion of the present state of the art of this device is given in Reference 14.

27.1 Basic Principles of Crossed-Field Tubes

27.1.1 Space Harmonics

As in the case of the O-type BWO discussed in Chapter 26, space harmonics play a vital role in the operation of crossed-field voltage-tunable oscillators and amplifiers. In the cases of the voltage-tunable magnetron, the BWO, and the backward-wave amplifier, there must be an electron beam whose velocity can be varied electronically and which is synchronous with a traveling-space harmonic on a slow-wave structure. Since the phase velocity of the space harmonic is a function of frequency, the velocity of the electrons in the beam determines the oscillation or amplification frequency.

Space harmonics are caused by periodic spatial perturbations along a waveguide structure or resonant cavity. Suppose we have a parallel strip transmission line with a wave traveling from left to right with the speed of light, as shown in Figure 27-1a. Now let us bend it into a zigzag path form-

(A)   *(B)*
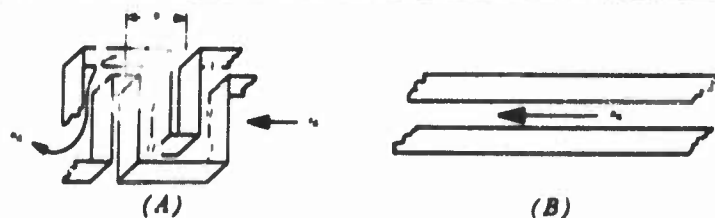
FIGURE 27-1   Propagation in the interdigital slow-wave structure.

ing 27-1b. The wave now travels along the path ABCDE, etc., with essentially the velocity of light. (It should be pointed out that the interdigital structure is only one type of slow-wave circuit, but is one of the simplest to describe and so will be used throughout this chapter for illustrative purposes.)

Now suppose an electron travels along the edge of the interdigital slow-wave structure from left to right as shown in Figure 27-1b. The electron is said to be traveling synchronously with a space harmonic if it always ex-

periences the same force or force-time variation every time it goes by a gap. In order for the electron to give up energy to the slow-wave structure, this must always be a net retarding force. Let us suppose that it takes the wave a time T_w to travel from E to E around the path $BCDE$. The apparent distance from B to E along the edge of the structure and the path of the electron stream is d. Now, since the field is reversed from one gap to the next, the electron must go from the middle of gap B to the middle of gap E in a half a period, $T/2$, (or any odd number of half periods), plus T_w. Expressing this mathematically:

$$v_{e_n}{}^+ = \frac{d}{(nT/2) + T_w} = \frac{2d/}{n + (2d//v'_w)} \qquad (27\text{-}1)$$

where d is the distance between finger centers, n is any odd integer, v'_w is the velocity with which the wave propagates down the circuit, and is equal to the phase velocity of the wave in the zigzag line, v_w, times the ratio of distances $BE/BCDE$. (This is approximately equal to the velocity of light times $BE/BCDE$).

Now, when the wave travels in the same direction as the electron and the electron keeps in synchronism with the fields in the gaps, as shown in Figure 27-1b, we say that the electron is interacting with a forward wave, and under such conditions we could have forward-wave amplification. The plus sign superscript on the symbol for electron velocity indicates that there is forward interaction. It is important to note that the electron need not travel with the progressive velocity of the wave down the circuit v'_w, as it keeps in synchronism with the fields in the gaps, but instead it can have a number of "synchronous velocities" called space-harmonic phase velocities corresponding to odd values of n of Equation 27-1. Because the electron travels alternately in front of a conducting finger, where it sees little or no longitudinal component of electric field, and then in front of a gap, where it sees the field in a particular phase, the field picture as seen by the electron is much the same as the picture seen by the human eye when viewing a motion picture, because of the shutter effect. A rotating wheel with spokes, in a motion picture, can appear to be rotating forward or backward, because of the shutter effect. In much the same way, because of the shutter effect, the fields in the gaps as seen by the electrons appear to move forward or backward with various apparent velocities for any given progressive velocity of the wave down the slow-wave structure. Let us now assume that the electron of Figure 27-1b is traveling from right to left while the wave is traveling from left to right. Now, in order for the electron to stay in synchronism with the fields in the gaps, the electron must travel from the middle of the gap at E to the middle of the gap at B in a half period (or any odd number of half periods) *minus*

T'_w. Thus the velocity of the electrons for backward-wave interaction is

$$v_{r_n}{}^- = \frac{2dj}{n - (2dj/v'_w)} \qquad (27\text{-}2)$$

where n is, again, an odd integer and the minus superscript indicates backward-wave interaction. This is the type of interaction we have in Carcinotrons, or BWO's.

In the preceding paragraphs we have been examining the apparent velocity with which the fields at the edge of the fingers of the slow-wave structure appear to the electrons. We could form a different viewpoint by synthesizing the known fields with a series of traveling-wave modal components. Were we to do this rigorously we would find that their velocities would be identical to those given in Equations 27-1 and 27-2. These traveling-wave modal components are called space harmonics, and for any given wave traveling down the slow-wave structure there are an infinite number corresponding to odd integers, n, traveling with various velocities both forward and backward.

27.1.2 Electron Interaction With the Space Harmonics

If an electron travels in the same direction and at the same velocity as a space harmonic, then it sees a steady force due to it, but it sees alternating forces due to all the other space harmonics. In most cases we can assume that the effects of these other nonsynchronous space harmonics on the electrons are small since the average of the forces taken over a long time is nil. Therefore, to the synchronous electron the space-harmonic field with which it is synchronous appears to be a static field. The typical shape of space-harmonic electric field for an M-type BWO, or Carcinotron, is shown in Figure 27-2.

In this view we have taken a cross section of the slow-wave structure perpendicular to the fingers of Figure 27-1b and have added a conducting plane parallel to the face of the fingers. This plane is called the sole; it has a high negative voltage relative to the slow-wave structure applied to it so as to establish a strong x-directed electric field shown in the figure as E_{dn}. Perpendicular to this in the negative y direction is imposed a strong uniform magnetic field. The spacing between the sole and the slow-wave structure is called a.

Now we will invoke a basic concept of electron motion in crossed electric and magnetic fields: in a region of uniform magnetic field and nearly uniform electric field in which the electric field is always perpendicular to the magnetic field, *electron always drift in the direction* $E \times B$ *and with the velocity* E/B, provided the time or space rate of change of the electric field is small compared with the length or period of a cycloid. Therefore a ribbon
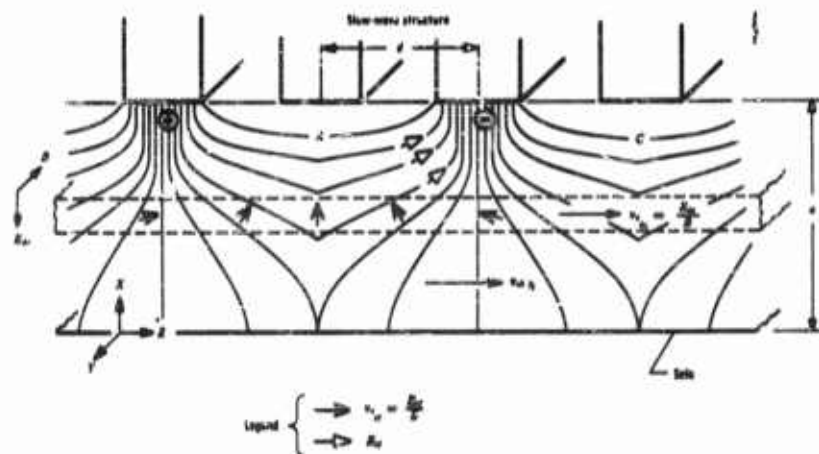
FIGURE 27-2   Electron interaction in the crossed-field tube.

of electrons as represented by the dotted lines of Figure 27-2 which are injected into the interaction space by an electron gun will drift in the z direction with the velocity E_{do}/B; and if the velocity of the space-harmonic fields is equal to this, then the moving electrons effectively see the space harmonic as static fields. Therefore whenever synchronism exists, electrons drift in the direction perpendicular to the space-harmonic fields as shown by the v-shaped arrows $(v_{e_{rf}})$, so that they congregate in the region A, forming one bunch per space-harmonic wavelength. In these bunches the electrons work their way across from the original beam position to the slow-wave structure, staying locked in synchronism as they do, and they move against the z component of the space-harmonic electric field, so that they give up their d-c acquired energy to the rf fields. Therefore energy is given up from the d-c applied field to the wave which gave rise to the space harmonic, causing the wave to grow. Slater (Reference 3) give a good discussion of crossed-field interaction for the magnetron oscillator.

Now in Figure 27-2, with the space harmonic and electron velocities being identical and from left to right, as shown by the arrows labeled v_{sh} and $v_{v_{do}}$, the wave which gave rise to this space harmonic could have been traveling on the slow-wave structure in any one of the three directions: (a) the positive z direction, (b) the negative z direction, or (c) the x direction with no propagation along z at all. Case a is the type of interaction in forward-wave amplifiers where the beam is progressively bunched as the wave grows; Case b is the type of interaction in backward-wave oscillators and amplifiers where the beam is progressively bunched as it progressively moves

toward the weaker fields, but regenerates itself because the bunched electrons feed energy to the low power or input end of the slow-wave structure; and c is the type of interaction often used in the voltage-tunable magnetron.

27.2 Description of Tube Operation

Let us look at the manner in which crossed-field interaction is used in the operation of the forward-wave amplifier, the backward-wave amplifier, the BWO, and the voltage-tunable magnetron. The schematic representations of these tubes are shown in Figure 27-3.

(a) Forward-wave amplifier

(b) Backward-wave amplifier

(or locked BWO)

(c) Backward-wave oscillator

(d) Voltage-tunable magnetron
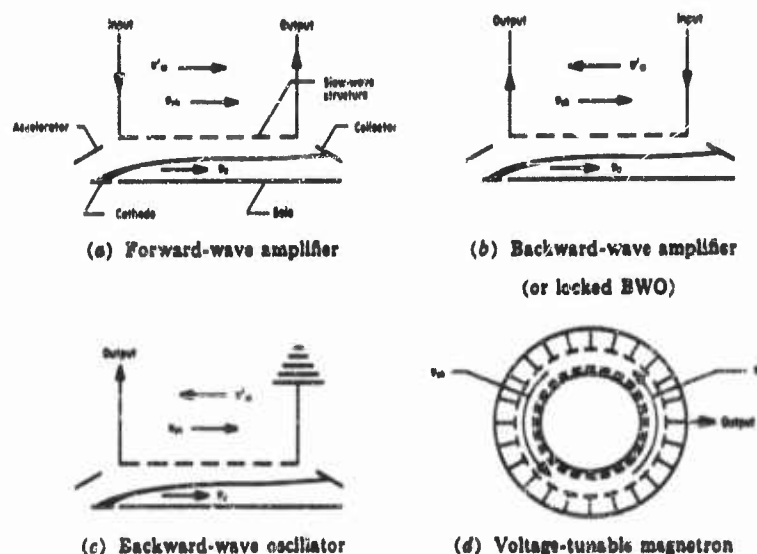
FIGURE 27-3 Principles of operation of four types of crossed-field tubes.

27.2.1 Forward-Wave Amplification

The input rf power of the forward-wave amplifier in Figure 27-3a is fed into the slow-wave structure at the end where the beam is injected, and the beam interacts with a synchronous forward space harmonic causing the signal to be amplified. As electrons interact they are drawn over to the slow-wave structure, and any electrons that have not completely interacted are collected on the collector. The propagation characteristic of the slow-wave structure is adjusted so that the first forward space-harmonic velocity is constant over a relatively wide band of frequencies, so that the amplification bandwidth is broad—typically as broad as a traveling-wave tube. It is convenient to express graphically the relationships between frequency and the

phase velocities of the space harmonics by means of the so-called ω-β plots.
(An alternative method is the use of the *c/v versus* λ diagram, which is
preferred by some workers in the field. This diagram has the following ad-
vantages: the lines of constant phase velocity are simply horizontal lines
corresponding to given values of the ordinate c/v; the group velocity is given
by the ordinate at the intercept of a line tangent to the phase-velocity char-
acteristic at any point; also the "cyclotron-resonance" relations are represented
by simple straight lines on this diagram). The value of β is defined as the ratio
of ω to phase velocity. Therefore in the case of space harmonics, β_a is equal
to $\omega/v_{e_n}{}^+$ for forward space harmonics, and $\omega/v_{e_n}{}^-$ for backward space har-
monics. Thus if we plot ω *versus* β for any given slow-wave circuit, the slope
of the line from the origin ($\omega = \beta = 0$) to the point of interest on the curve
is equal to the space-harmonic velocity. The velocity with which energy or
modulation intelligence propagates down the line is called the group velocity,
and can be shown to be equal to $\partial\omega/\partial\beta$. Therefore the slope of the ω-β curve
at any point is equal to the group velocity. A suitable ω-β plot for the
forward-wave amplifier is shown in Figure 27-4, with the solid segments

FIGURE 27-4 ω-β plot showing for-
ward-wave and backward-wave inter-
action.

of the curve corresponding to the for-
ward space harmonics and the dashed
segments corresponding to the back-
ward space harmonics. It may be seen
that if the electrons have a velocity
corresponding to the slope of the line
oab they are synchronous with the space
harmonic over the frequency range ω_a
to ω_b, and therefore this is the amplifi-
cation bandwidth. The shaping of the
ω-β curve for synchronization of the

electrons with the space harmonic over the range ω_a to ω_b is called "broad-
banding" the structure. The low- and high-frequency cutoffs, ω_{l_1} and
ω_{l_2}, define the first passband. It should be pointed out that there are many
higher-frequency passbands separated by stop bands, with ω-β characteristics
similar to that of the first passband. However, their fields are usually so
weak as to be insignificant in comparison with the fields of the first pass-
band, so they usually do not cause trouble.

It is to be noted that there is no inherent feedback mechanism in the for-
ward-wave amplifier as there is in the backward-wave amplifier to be de-
scribed next.

27.2.2 Backward-Wave Amplifier
In the backward-wave amplifier shown in Figure 27-3b, the power is fed

Into the collector end of the beam and the electrons interact with a synchronous backward space harmonic of this wave. Therefore, the wave grows as it travels from right to left. It is to be noted that the maximum beam bunching fields exist (if complete interaction of all the beam has not occurred) at the input end of this tube, and these bunches induce some signal at the input end of the structure which reinforces the input signal, constituting an inherent regenerative feedback mechanism. The gain of this device increases as the beam current is raised, and because of the regenerative effect, the gain becomes infinite at a particular beam current called the oscillation starting current, I_s. In order for the tube operation above this current to be controlled by the input signal, it must be strong enough to override the regenerative effect of the beam feedback; otherwise the tube oscillates independently. At beam currents below the oscillation starting current, the tube is a voltage-tunable regenerative-feedback amplifier. At beam currents above the oscillation starting current the tube is a voltage-tunable locked oscillator in the presence of sufficient input rf driving power, and in the absence of input power, the tube becomes a voltage-tunable oscillator, or an M-type Carcinotron. The portion of the ω-β plot used in a backward-wave amplifier or oscillator is shown in Figure 27-4 as the segment of a dashed curve from c to d. In this case only the backward space harmonic is used. As may be seen, the backward space-harmonic velocity increases as one goes from point c on the curve to point d, and the frequency also increases. Therefore, to voltage tune a backward-wave amplifier or oscillator, one merely varies the electron velocity, which must be synchronous with the backward space-harmonic velocity. In most backward-wave amplifiers or oscillators the forward-wave portion does not need to be "broadbanded" as in Figure 27-4,

and usually the ω-β curve looks typically like that shown in Figure 27-5. The bandwidth used is typicallly that between ω_o and ω_d which is about a 1.5:1 frequency range.
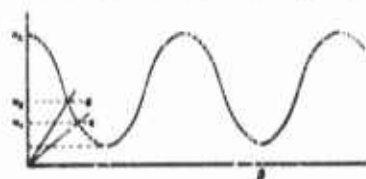
It is to be noted that, at any given electron beam velocity, or voltage, the amplification bandwidth of the backward-wave amplifier exists only in the region where the line through the origin crosses the backward-wave portion of the ω-β characteristic, and is relatively small (1.05:1) as compared with a forward-wave amplifier or traveling-wave tube (up to 2:1); but this narrow bandwidth can be voltage tuned over a 1.5:1 frequency range.

FIGURE 27-5 ω-β plot for a backward-wave amplifier or oscillator.

Let us now consider operation of the backward-wave amplifier below starting current. Let a backward-wave tube be outfitted with an external rf con-

nection at both the gun end (output) and the collector end (input) of the delay line. Then if an external signal is injected at the input, one observes a backward gain (defined as 20 times the log₁₀ ratio of output rf voltage to input rf voltage) under conditions of velocity synchronism between the beam and the space-harmonic wave (interacting fundamental spatial harmonic) at operating currents below the starting current, I_s. This backward gain attains very high values at beam currents just below I_s and becomes infinite when the beam current becomes equal to I_s. As discussed earlier, a type of positive feedback is inherent in backward-wave interaction and, therefore, the backward-wave amplifier is regenerative in nature.

Small signal theory indicates that in order to attain a gain of greater than 20 db, for most practical tubes, the operating current must be only about 2 or 3 percent below the starting current. Since the gain theoretically varies as $1/I - I_s$, at current approaching I_s, it is clear that a high degree of current stability is necessary in working with backward-wave amplification. The gain drops off very rapidly as the input frequency or anode voltage V_s is varied from the value for synchronism. In the region of high gain, as the current is changed the gain and the bandwidth change in such a manner as to keep the gain-bandwidth product approximately constant. For high values of gain on the order of 20 to 40 db, the theoretical bandwidth between 3-db points is very small, of the order of 0.01 to 0.1 percent. For equivalent interaction parameters, the more dispersive the delay line, the smaller the bandwidth at a given gain.

The center frequency of the amplification band is voltage tunable—usually by variation of the anode voltage V_s. For constant gain, however, one must vary the operating current with the voltage since the starting current varies with frequency or voltage.

A greater gain-bandwidth product, resulting in less critical dependence on operating current, can be obtained by the use of two circuit sections in one tube, as in the Cascade (References 15 and 16) backward-wave amplifier. The improved operation is obtained at the expense of increased complexity of the tube design and is limited by saturation effects (Reference 16) peculiar to crossed-field interaction.

Let us now consider the backward-wave amplifier in superregenerative operation. In the operation of a high-gain backward-wave amplifier, the current is close to the starting current where the gain varies very rapidly with current. Therefore, one may expect extreme operational instability which limits the practical usefulness of such operation. To avoid this instability, one may utilize superregenerative amplification.

In a superregenerative amplifier the state of operation is alternated between a nonoscillating and an oscillating condition. In the backward-wave amplifier this could be done by varying the current cyclically between a

current below I_s and a current above I_s. The frequency of this alternation should be much greater than that of the highest frequency of the modulation to be applied to the input signal. This alternation is called the quenching cycle. In the ideal quenching cycle of such an amplifier, the tube is in a non-oscillating or quenched state during a small part of the quenching cycle; and during most of the cycle the tube is placed into conditions under which oscillations may build up. In this latter part of the cycle, a regenerative or "sensitive" state prevails.

The operation of the superregenerative amplifier utilizes the dependence of the rate of oscillation buildup on the amplitude of an input signal present during the sensitive part of the quenching cycle. The nature of the amplifier with regard to the variation of gain with input signal depends upon the period of the quenching cycle compared to the normal buildup time of oscillations. If the quenching frequency is very high so that the oscillation never reaches a saturated state during the sensitive part of the cycle, then the gain is constant over a great range of input power, and linear amplification results. For lower quenching frequencies such that the oscillation saturates during the sensitive part of the cycle, the gain decreases with input power and the amplification tends to be logarithmic.

A knowledge of the buildup time in the absence of an input signal is important for the understanding and design of superregenerative amplifiers. A theoretical determination of the rate of buildup including the effects of space charge and line attenuation is complicated, but if one neglects these effects and assumes the oscillation builds up as $e^{t/\tau}$ in a small-signal state, then it is found that τ decreases rapidly from infinity for I just greater than I_s and decreases slowly for $I > 3I_s$. This calculation assumes that the current is instantaneously set to the value I at $t = 0$, whereas in reality the current rises to the value of I in a finite time after passing I_s. From τ one can estimate the total buildup time from noise to usual operating levels. For $I/I_s > 1.5$ one finds a theoretical buildup time of the order of 10^{-7} to 10^{-8} seconds for most practical tubes. The effect of space-charge is to decrease this time so that these values may be considered as upper limits. By operating closer to I_s one may increase the buildup time considerably and therefore the gain, but this increase is limited by operational instability as I is made too close to I_s, where τ varies rapidly with I. The choice of I is governed by the requirement of a buildup time sufficiently greater than the required quench cycle so as to obtain sufficient gain.

In utilizing a backward-wave amplifier for superregenerative operation over a wide range of signal frequencies it is necessary, as in operation below starting current, to vary the operating current with frequency in order to maintain a constant gain since the starting current varies with frequency.

The superregenerative amplifier has a gain-bandwidth product which is

much greater than that of the regenerative amplifier. The bandwidth will vary with the quench frequency and with the quench waveform. Generally, the less time spent in the regenerative or sensitive part of the quench cycle the wider the bandwidth.

27.2.3 Backward-Wave Oscillation

Figure 27-3c shows a schematic representation of the M-type BWO, or Carcinotron, which differs from the backward-wave amplifier only in that the input is terminated in a reflectionless load, and the tube is always operated above the oscillation starting current, I_s. The ω-β characteristic of this type of tube is shown in Figure 27-5. The rf power output appears at the output terminal, but, because of incidental reflections along the slow-wave structure or at the output, some power is reflected and progresses down the line toward the collector end of the tube. There are normally no space harmonics interacting with this reflected power, but if the power is re-reflected it combines with the electronic feedback mechanism both regeneratively and degeneratively, causing periodic wiggles or holes in the voltage-frequency curve and power-frequency characteristic. Therefore, the slow-wave circuit at the collector end is terminated as reflectionlessly as possible.

The frequency of oscillation is determined by the velocity synchronism condition between the beam and circuit space harmonic. Variation of electron velocity v_e, which is equal to $(V_{kl} + V_{ks})/Ba$, voltage tunes the tube. (See Figure 25-13 for definition of voltages.) The electrical parameter of the tube which is varied to accomplish voltage tuning is either V_{kl} or V_{sk}, the cathode-line or sole-cathode voltage. For sweeping the complete band of the tube, V_{kl} is varied; while for tuning or frequency modulation over a narrower band the sole voltage V_{ks} is varied.

The beam current may be controlled by either the grid voltage V_{kg} or the accelerator voltage V_{kn}. By these means, amplitude modulation may be accomplished.

The efficiency and power rapidly increases with current for $I/I_s > 1.5$, and the efficiency tends to level off for $I/I_s > 3$. The operating currents, and therefore the output power, are limited by the delay-line dissipation and usually are in the range $2I_s$ to $5I_s$. Moderate to high efficiencies, 15–50 percent, inherent in M-type devices, have been attained in CW backward-wave oscillators of the M-type in the microwave range of 1000–10,000 Mcs.

If the attenuation region in the tube is perfect then there is no frequency pulling, in theory. In this case a Rieke diagram of the tube should show circles of constant power but no frequency shifts. If, in addition, the output transition presents no reflection, then the circles of constant power are

centered at the origin. This shows that the maximum useful power is obtained in the load when the tube operates into a matched load as seen just inside the output transition. With a perfect attenuation region the tube can work into a load with any value of VSWR without frequency pulling or frequency discontinuities as occur, for example, in the "sink" region of the Rieke diagram for a magnetron.

An imperfect termination, or irregularities in the delay-line structure, leads to the introduction of frequency pulling and the possibility of frequency discontinuities for loads of high VSWR. This is because the presence of reflections at both ends of the delay line permits an additional type of feedback which may be regenerative or degenerative depending on the electrical length between reflections. Whenever a parameter such as the anode voltage or output load phase is varied through a condition where degenerative feedback occurs due to reflections, frequency discontinuities or holes may result. For loads of high VSWR and a poor attenuation region, ten or more frequency holes may result. Well-designed attenuation regions and accurate delay-lines will eliminate this type of frequency hole. The space charge of the beam has a second-order effect on frequency, causing frequency pushing. Pushing is defined as the change of frequency with beam current when all other parameters are held constant. Frequency deviations due to pushing are typically of the order of 0.6 percent of center frequency as the beam current is varied so as to cause the power output to vary from zero to its normal operating value.

If the backward-wave tube is provided with an input as shown in Figure 27-3b, then the injection of a signal at the input may result in locking of the free-running oscillation to the frequency of the injected signal. The theoretical determination of the locking range of frequency is complicated if space-charge effects are included. A classical small-signal treatment which neglects the space charge shows that the locking range of frequency is proportional to the ratio of injected power to the output power of the free-running oscillator. When this ratio is unity, theory predicts a typical locking range of a few percent. Experimentally, one often finds locking-range values greater than those predicted by theory.

If one injects power of the same order as the output power of the free-running oscillator there appears a new feature in the performance more significant than the mere frequency locking. With the optimum condition of frequency locking, the total output power of the locked oscillator can be greater than the sum of the injected power and the free-running output power. This means that the efficiency of the locked oscillator, defined as $(P_{out} - P_{in})/P_{d \cdot c}$, is greater under the locked conditions than in the free running state. This difference can approach a ratio of 2, which indicates that

a cascade arrangement of two oscillators can deliver more power at higher
efficiency than simple parallel arrangements of the oscillators. The driven
tube in such a cascade arrangement is called a Bitermitron,* and the com-
plete system of the two or more tubes is called a Bitermitron system.

A qualitative explanation for the difference in efficiency lies in the fact that
in the free-running case the rf field near the input is small according to the
usual field distribution in backward-wave oscillators; but in the driven case,
the rf field is large near the input. Therefore, one expects the interaction in
the region near the collector to be more efficient in the driven case.

27.2.4 Voltage Tuning of the Magnetron Oscillator

The voltage-tunable magnetron is different in several ways from the tubes
previously discussed. First of all, the beam and slow-wave structure may
both be re-entrant. Second, the rf beam impedance is decreased and the d-c
anode current is controlled and limited either by means of temperature limit-
ing the emission from the cathode or by the use of an injected beam. In the
latter case a whirling tubular sheath of electrons from a cathode is injected
into the generally cylindrical re-entrant interaction space formed by the
anode and an inner nonemitting electrode, often called the sole, and since
the rate of electron injection into the interaction space may be controlled,
the apparent emission will appear to be temperature limited.

Since the slow-wave structure is electromagnetically re-entrant in the
typical case the operating values of β are fixed at those discreet values, or
modes, which satisfy the re-entrance conditions, generally an integral number
of wavelengths around the structure.

The slow-wave structure then forms an integral part of an rf resonant
circuit, and this is then heavily resistively loaded by the output line, or by
other means. The electrons which form the rotating sheath are bunched into
"spokes" as in the case of the conventional magnetron oscillator, and be-
cause the effective emission is limited the rotational velocity of these bunches
can be varied by varying the sole-anode voltage, without affecting the amount
of current drawn by the anode.

These electron bunches induce inductive or capacitive currents in the slow-
wave structure, depending on whether they travel slower or faster than the
speed corresponding to the midband frequency of the tank circuit. This, in
effect, moves a small portion of the ω-β curve up and down in the vicinity
of the fixed value of β corresponding to the operating mode of the slow-
wave structure, as illustrated by Figure 27-6. A voltage-tuning diagram can

*Raytheon trade name.
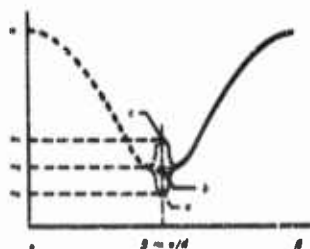
FIGURE 27-6 ω-β plot for a voltage-tunable magnetron.

be obtained by following points a, b, and c, on the ω-β curves, which correspond to sole-anode voltages V_a, V_b, and V_c. The typical voltage-frequency variation of the voltage-tunable magnetron is linear and proportional.

It should be pointed out that the value of β can, in general, be fixed at as many discreet values form 0 to $2\pi/d$ as there are anode segments in the slow-wave structure and that voltage-tunable operation may be obtained in either forward-wave, backward-wave, or standing-wave modes.

27.3 The M-Type Backward-Wave Oscillator (M-Type Carcinotron)

27.3.1 Physical Characteristics

The basic elements of a backward-wave tube are indicated schematically in Figure 27-7 with the usual arrangement of applied d-c voltages.
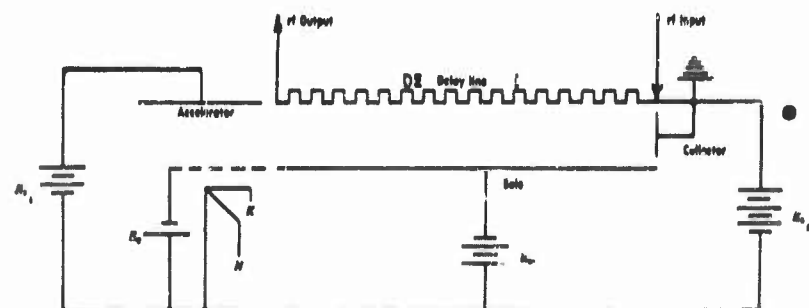
FIGURE 27-7 Schematic diagram of an M-type BWO.

The arrangement of the basic elements may be separated into three regions: the gun region bounded by the cathode, grid, and accelerator; the interaction space bounded by the delay line and the sole; and the collector region usually bounded by the actual collector element and the sole. In practical tubes, the collector, as well as the delay line, is an integral part of the tube body which is maintained at ground potential.

With the appropriate values of applied potentials and the appropriate value and direction of magnetic field (see Figure 27-7), the electron beam is formed in the gun region, passes through the interaction region, and is collected under operating conditions at both the delay line and the collector. In

order to obtain the correct beam transmission, it is important to apply the magnetic field in the correct sense.

In most tubes, the delay line is of the interdigital type. In backward-wave oscillators, the rf input is absent and a portion of the delay line adjacent to the collector is modified suitably with means of attenuation so as to provide an adequate termination for the absorption of any power reflected from the output. In the amplifier or Bitermitron, no attenuation is specifically introduced into the delay line and an rf input connection is provided.

In most practical tubes the interaction space is annular rather than linear. Figure 27-8 shows a cross section of a typical tube, and, pictorially, the path of the electron beam and the rf energy under oscillating conditions.
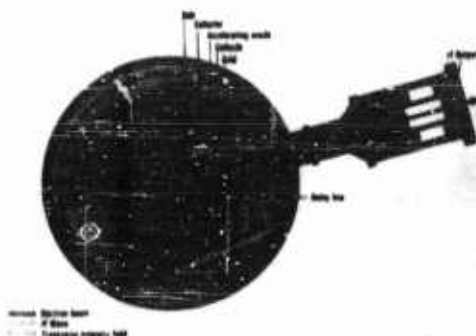
FIGURE 27-8 Elements of an M-type BWO.

FIGURE 27-9   S-band M-type BWO showing internal construction.

The header is that portion of the tube in which the electrode leads are brought out to external terminals; it comprises the principal projection from the main cylindrical body of the tube in addition to the rf output. Both radial and axial headers have been used, but the axial is preferred because of advantage in tube construction and because it is more adaptable to the use of a toroidal-type magnet, which has advantage over other types of magnets, as detailed below. Figure 27-9 shows a packaged S-band tube with a part of the tube cut away to show some of the internal construction. The form of the toroidal-type magnet is also shown. Its self-enclosure causes its field to decrease very rapidly outside the magnet; this valuable shielding function is not found in other types of magnets. For example, for a typical S-band tube utilizing a toroidal magnet, the field at any point 5 inches away from the external boundary is less than 50 gauss when the magnet is stabilized at a gap field of 1200 gauss. In general, the field outside the magnet decreases as

$1/r^2$, where r is the radius from the center of the magnet. This dependence on r is simply that of a dipole.

Figure 27-9 also shows in cross section a cylindrical liquid-cooling channel at the periphery of the tube body and one of the external connections to this channel. The liquid cooling of the tube cylinder removes the heat which is generated by the interception of the beam at the delay line and the collector. In most tubes designed for applications where thermal frequency drift requirements are not severe, the sole is not specifically designed for direct cooling; but some cooling of the sole is effected by cooling of the sole support-header projection.

The inside of the body of a typical C-band tube is shown in Figure 27-10, showing the electron gun end of the interdigital slow-wave structure. Figure 27-11 shows the collector end of the slow-wave structure. Reflectionless ter-

FIGURE 27-10 Photograph of the electron-gun end of a slow-wave structure.

FIGURE 27-11 Photograph of the collector end of a slow-wave structure.

mination of the structure is achieved by putting a thin coating of iron or other lossy material on a number of the end fingers. Lossy ceramic terminations are sometimes used in low-power tubes, and at low frequencies (below L-band) this end of the slow-wave structure is sometimes brought out of the tube through another terminal which is terminated externally. The collector is located beyond the end of the rf termination of the slow-wave structure, and consists of a beam-interception structure which has a good thermal path to the body and cooling jacket of the tube so that it can readily dissipate the beam power.

The radial d-c electric field is provided by the sole electrode, which is

placed inside the slow-wave structure. The sole is often used also for mounting the electron gun. A photograph of the inside of an M-type BWO with the sole and gun mounted in position is shown in Figure 27-12. Because the sole must operate at a high negative voltage relative to the body, it is mounted on an insulating ceramic. The electron gun electrodes are also mounted with insulating ceramics.

A cross-section view of the electron gun typically used in magnetron type backward-wave oscillators and amplifiers is shown in Figure 27-13. This is a
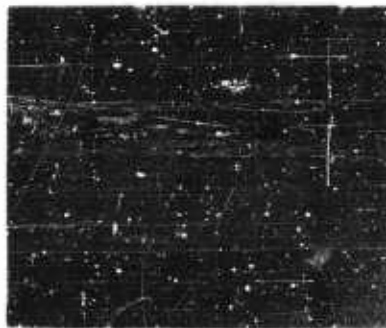
FIGURE 27-12 Photograph of the inside of an M-type BWO showing sole and gun mounted in position.
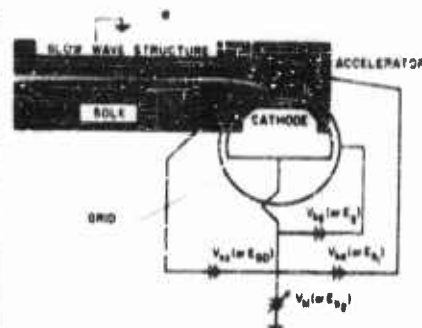
FIGURE 27-13 An electron gun of an M-type BWO showing the beam formation.

photograph of the actual beam in an M-type BWO gun as studied in a manned space chamber. The electrode cross sections have been emphasized and a schematic external circuit diagram has been added to the photograph. In this gun the electrons, emitted from the strip cathode, are initially accelerated toward the accelerator because of the positive voltage impressed on it relative to the cathode. The magnetic field bends the electron trajectories to the left, forming a half a cycloid; and then the electrons enter the slow-wave structure space with its stronger electric field, which straightens out the trajectories. The beam current is controlled either by varying the positive accelerator voltage or by varying the negative voltage applied to the channel-shaped grid. It is important to note that neither of these electrodes has a first-order effect on the electron velocity in the sole-circuit region since the velocity there depends on E/B and not on the entrance conditions. The electron gun is designed to give the thinnest and smoothest beam possible, even though the sole-to-line voltage varies by a factor of $2:1$.

The external appearance of the M-type BWO roughly resembles that of a CW magnetron, except for the absence of any mechanical-tuner mechanism. Figure 27-14 shows typical 150-watt tubes in the C- to X-band frequency
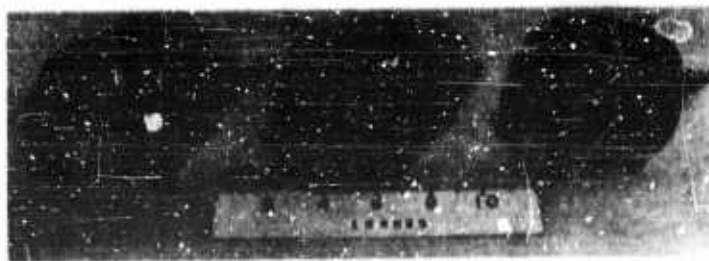
FIGURE 27-14 Typical 150-watt CW M-type BWO's covering the frequency range of 4800 to 11,000 Mc.

FIGURE 27-15 Family of 200-watt CW M-type BWO's.
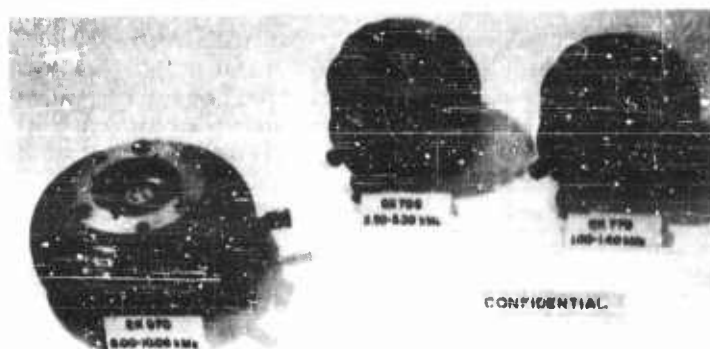
FIGURE 27-16 Typical 400-watt CW M-type BWO's.

range. Each has a packaged weight of 15 lb with a magnet diameter of $6\frac{1}{2}$ in. and an over-all axial dimension of $4\frac{1}{2}$ in. Figure 27-15 shows a family of 200-watt oscillators ranging from an L-band tube to an X-band tube. These tubes generally weigh about 20 lb and have a magnet diameter of $7\frac{1}{4}$ in. The physical uniformity in this family of tubes is desirable and feasible in the principal microwave range of 1 to 11 kmc. In Figure 27-16 are shown some typical 400-watt oscillators, which are very close in size and weight to the 200-watt tubes.

27.3.2 Operating Characteristics

The principle feature of BWO's is their voltage-tuning capability. In the M-type BWO, voltage tuning is obtained by varying the line-to-sole voltage $V_{kl} + V_{ks}$ to which the beam velocity is proportional. By equating the beam velocity as a function of voltage to the expression (Equation 27-2) for circuit phase velocity, a frequency-voltage relation is derived:

$$V_{kl} + V_{ks} = \frac{k_1 f}{1 - k_2 f} \tag{27-3}$$

The second term in the denominator of Equation 27-2 is typically of the order of $\frac{1}{3}$ to $\frac{1}{5}$ in the voltage-tuning range of most tubes so that a variation of 2:1 in $V_{kl} + V_{ks}$ results in about a 1.5:1 variation in frequency. In existing tubes, this is about the maximum tuning range consistent with good efficiency.

It is important to keep the electron beam in as nearly the same position as possible as the tube is voltage tuned. The most satisfactory method of voltage tuning from the standpoint of optimum efficiency and tuning range is to fix the voltages from cathode to the accelerator, sole, and grid, and vary the anode voltage V_{kl} from this combination to ground. Voltage-tuning ranges of 1.5:1 can be obtained in this way with good power output, but they have the disadvantage of requiring the modulating circuit to carry the beam current. Also, the sole, grid, and accelerator voltage-supply chassis must be varied in voltage with respect to ground, and this puts a practical limit on the rate at which the frequency can be varied. From the circuit standpoint, the most satisfactory way to voltage tune the tube is to vary the sole voltage V_{sk} only, leaving all other voltages fixed. Since the tube can be designed so that the sole current is very small, this method of tuning obviously eliminates many circuit problems. The widest practical tuning range that has been achieved to date by this method is about 1.2:1, limited by deterioration of electron optics at the end of this tuning range. Tubes can be designed to cover both the upper and lower 1.2:1 portions of the total 1.4:1 range by two choices of anode voltage.

FIG. 9A.19
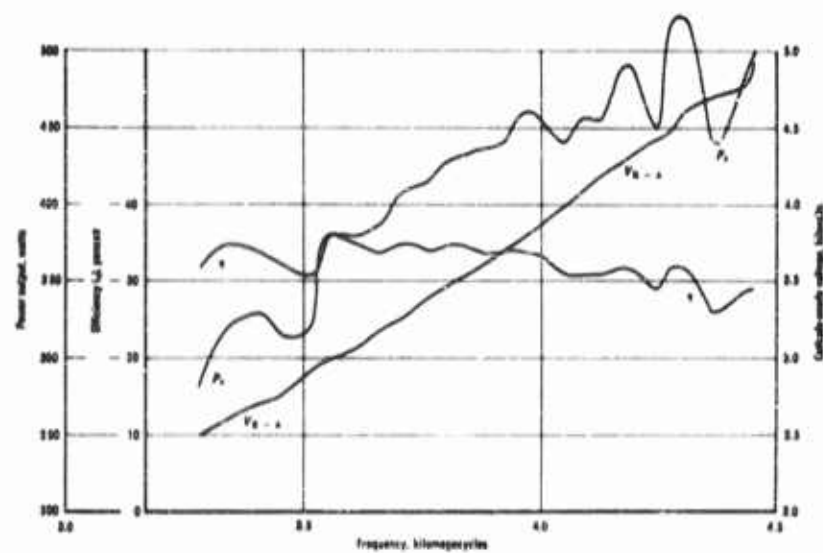
FIGURE 27-17 Performance curves for a 200-watt S-band M-type BWO with anode voltage tuning, BWM-7097. $I_b = 300$ ma; $R_{aa} = 800$ volts.
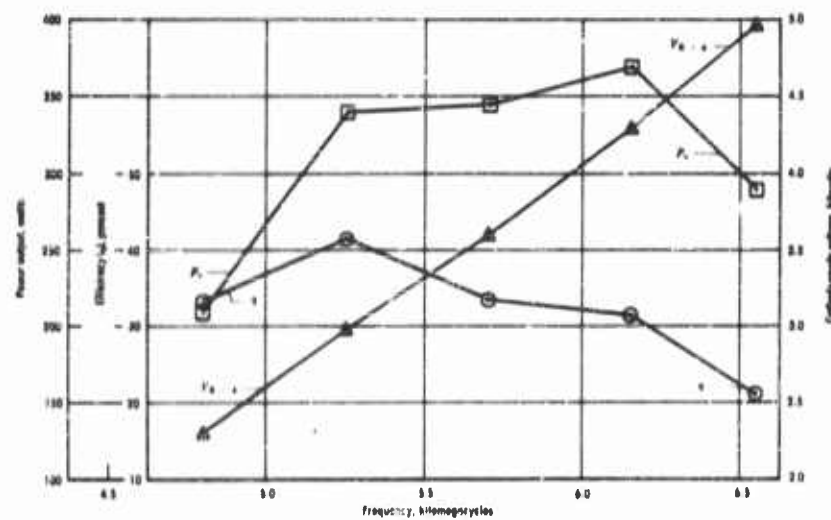
FIGURE 27-18 Performance curves for a 200-watt C-band M-type BWO with anode voltage tuning, L-3146. $I_b = 275$ ma; $R_{aa} = 1600$ volts.
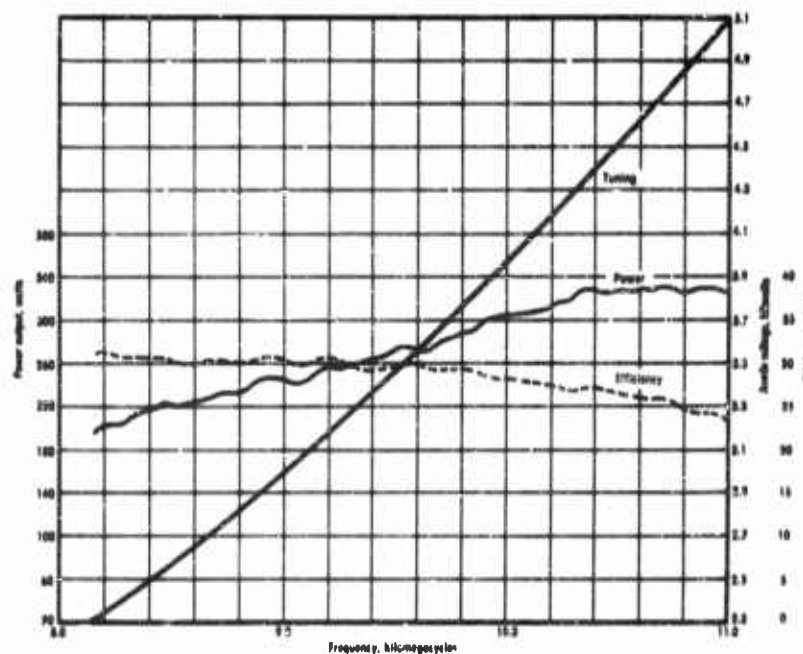
Figure 27-19 Performance curves for a 200-watt X-band M-type BWO with anode voltage tuning, QK634. $I_{sg} = 275$ ma; $K_{ss} = 1600$ volts.

Curves of frequency, power, and efficiency *versus* anode voltage for 200-watt CW tubes at S-, C-, and X-band frequency ranges are shown in Figures 27-17 through 27-19. In Figure 27-20 are shown similar curves for a 400-watt L-band oscillator. For these curves, the accelerator, grid, and sole voltages were held constant with respect to the cathode. For these and other existing tubes, the tuning ranges are typically 1.4:1. Sole-tuning curves for L- and C-band tubes are shown in Figures 27-21 and 27-22; it may be seen that the tuning ranges for existing tubes are about 1.2:1. The tuning rates, i.e., megacycles per volt, are essentially the same as in the anode-voltage tuning curves. In part, the high-frequency end of this sole-tuning range is limited by reduction of the efficiency and the low end is limited principally by excessive sole current. The fact that only the sole voltage needs to be varied, however, makes this type of operation attractive for equipment use in spite of the restricted tuning range.

In Table 27-I are listed all active tubes of the M-type BWO class as supplied by the Advisory Group on Electron Tubes (AGET), valid on June 1, 1959. This list includes all M-type BWO's in active development or in production
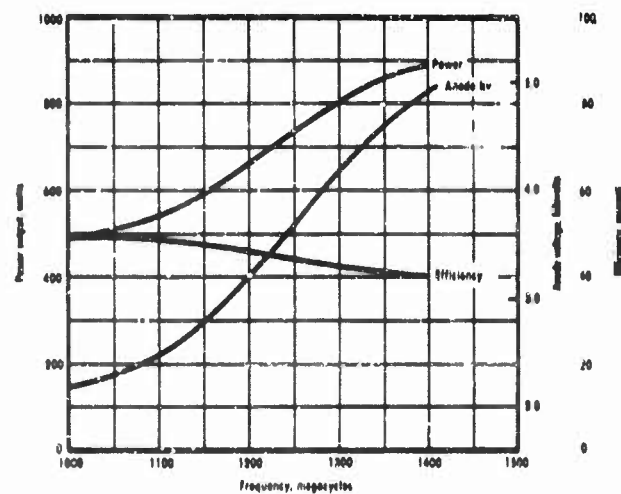
FIGURE 27-20 Performance curves for a 400-watt L-band M-type BWO with anode voltage tuning, QKA778. I_{bg} = 450 ma. E_{sg} = 1100 volts.
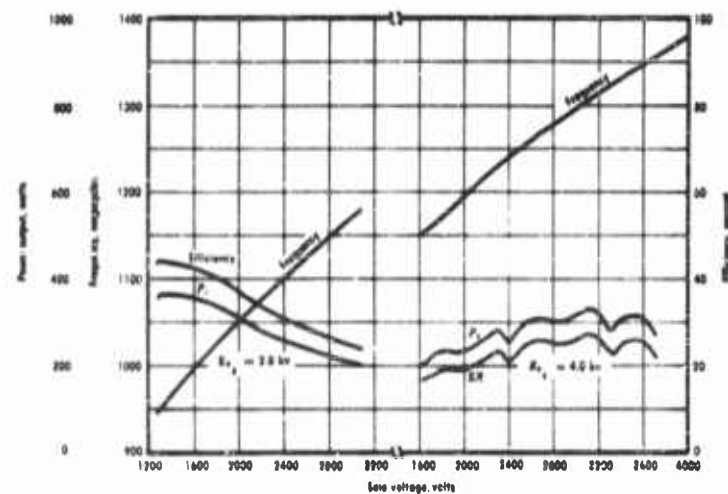
FIGURE 27-21 Performance curves for a 200-watt L-band M-type BWO with sole voltage tuning, QK656. I_{bg} = 300 ma.

TABLE 27-I. ACTIVE TUBES

TI No. or WE No.	Service	Cont. Mfr.	Tube Type	Function	Frequency Min. (mc)	Frequency Max. (mc)	Power Output CW	L. Wt. (lbs)	
MB080-04		LIT P	L-3241	OB	950	1310	150 w		
MB086-01	FW	RAW B	QK-656	OB	960	1310	300 w	5.3	Ruggedized
	Z	RAW C	QK-776	OB	1000	1400	400 w	5.0	
MW1814	FW	EVW B	CL-Series	OB	1150	1700	300 w	3.0	
MB080-05		LIT P	L-3142	OB	1350	1615	150 w		
MB086-02	FW	RAW A	QK-657	OB	1302	1850	300 w	5.3	Ruggedized
MW1814	FW	EVW B	CL-Series	OB	1540	1900	300 w		
MB080-06		LIT P	L-3143	OB	1300	2550	550 w		
MB086-03	FW	RAW A	QK-658	OB	1600	1850	300 w	5.3	Ruggedized
MB170-01	FW	BOM A	BL-692	OB	1550	3300	150 w	5.0	
MB176-01	FW	BOM A	BL-897	OB	1500	3300	150 w	5.0	Ceramic Seal BL-693
MW1615	FW	RAW C	QK-631	OB	1500	3300	200 w	5.2	Replace QK-658 Ruggedized Unclassified
	Z	RAW B	QKA700	OB	1500	3300	430 w	5.0	Unclassified
MW1880	Z	RAW B	QK-540/631	OBA	1500	3300	300 w	5.3	Matched OB AM Pair
MB144-01	FW	BOM A	BL-654	OBA	1500	3300	500 w		Modified BL-892 Pair Separate Magnet
MW3102	NB	RAW B	QK-618/625	OBA	1600	3300	1.4 kw	5.3	Matched OB AM Pair
MB080-03	FW	LIT P	L-3144	OB	1500	3450	150 w		17-20 lb Packaged Airborne or High-Alt. Use
MB086-04	Z	RAW B	QK-659	OB	3350	4450	150 w	5.0	
MB086-04	FW	RAW A	QK-733	OB	3150	4450	300 w	5.3	
MB170-01	AB	EVW A	BWM1046	OB	3350	4450	300 w	3.0	
MW3697	FW	RAW B	QK-670/659	OBA	3350	4450	800 w	5.0	Matched OB AM Pair
MB080-07		LIT F	L-3145	OB	3400	4550	150 w		
M5065-01	A	MIU F		OB	4000	6000	10 kw	5.0	Experimental Tube Separate Magnet
MB086-05	FW	RAW A	QK-660	OB	4400	6050	300 w	5.3	
MB080-02	NO	LIT A	L-3146	OR	4800	6550	150 w	5.0	
MB080-03	FW	LIT A	L-3717	OR	4900	6550	150 w	5.0	Gridded Forced-Air Cooling
MB086-06	Z	RAW A	QK-881	OR	6000	8300	300 w	5.3	Ruggedized
MW3717	RC	LIT C	L-3147	OB	6300	8550	150 w	5.0	
MB278-01	NB		P	OR	8000	10,000	1.0 kw		
MB110-15	NB	RAW A	QK-610/370	OB	8000	10,000	1.0 kw	5.0	Matched OB AM Pair
MW1301	NB	RAW B	QK-570	OB	8000	10,500	500 w	5.0	
MB086-07	FW	RAW C	QK-834	OR	8150	11,000	150 w	5.3	Prod refined QK-551
MW1888	FW	LIT C	L-3234	OR	8450	9650	125 w	AC	Forced-Air Cooling Narrowband L-3148
MW3705	FW	RAW B	QK-677/353	OBA	8600	9600	100 w	5.2	Matched OB AM Pair
MW1882	NA	LIT C	L-3148	OB	8800	11,600	125 w	5.0	
MB149-01	FW	VAR A	VA-166	OR	14,900	18,100	100 w	5.0	
MB153-01	FW	BOM A		OB	31,000	36,000	75 w		Planned
MX111-01	NB	RAW A	QK-719	OB	50,000	60,000	300 w	5.0	Feasibility Study
MB100-01	NC	VAR A	VA-165	OB	33,750	43,250	50 w		Separate Magnet 50-W Interim Goal
	FR	CNF A	CM-32	OB	200	290	5 kw		
	FR	CNF A	CM-34	OB	185	400	8 kw		
	FR	CNF A	CM-36	OB	390	570	23 kw		

NOTES

Security classification of all tubes is CONFIDENTIAL except where UNCLASSIFIED
is noted in "Remarks" column.

Structure of all tubes is backward wave, M-type.

All tubes are voltage tuned.

Cooling of all tubes is liquid except where "forced air" is noted in "Remarks" column.

Magnets are all permanent, integral, except where "separate" is noted in "Remarks"
column.

TI or VC Number: prefix indicates technical information category.

Sponsor:

  A: Army, agency unspecified
  AS: Army, Signal Supply Agency (USASSA)
  FW: Air Force, WADC
  NA: Navy, BUaer
  NO: Navy, BUord
  NS: Navy, BUships
  SC: Army, Signal Corps, USASRDL
  Z: Industry sponsored
  FR: Air Force, RADC

Company or Laboratory:

  BOM: Bomac Laboratories
  LIT: Litton Industries
  MIU: The University of Michigan
  RAW: Raytheon Company, Waltham
  SYW: Sylvania Electric Products Company, Woburn
  VAR: Varian Associates
  CSF: Compagnie Generale de T.S.F.

Status of Project:

  A: Active
  B: Completed
  C: Commercially announced or in production
  P: Production

Function:

  OS: Oscillator
  OSA: Amplifier-oscillator
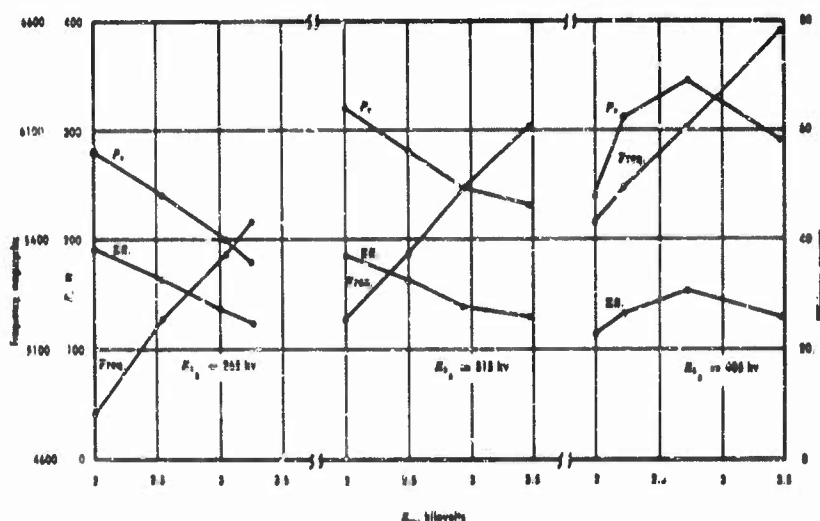
FIGURE 27-22 Performance curves for a 200-watt C-band M-type BWO with sole voltage tuning, LJ396. $I_b = 275$ ma.

In the United States. Many tube types in the principal microwave range of 1 to 11 kmc are shown to be in or close to production, whereas tube types in the K-band range are still under development and not yet in production.

The most widely used method of modulation of the M-type BWO is frequency modulation by varying the sole voltage. The frequency deviation obtainable is the same as the sole-tuning range, i.e., 1.2:1. The maximum rate at which this deviation can be swept before deterioration of the tube characteristics begins to occur is at least 30 Mc at L-band and 300 Mc at X-band. The chief practical problem in obtaining high frequency-modulation rates is in overcoming the effects of sole-to-ground capacitance. At X-band, this capacitance is typically 35–50 $\mu\mu f$, at S-band 60–70 $\mu\mu f$, and at L-band 80 $\mu\mu f$.

Amplitude modulation is achieved by varying the accelerator voltage or the grid voltage, thus varying the beam current. The power output may typically be varied from 0 to 3 times the nominal CW power of the tube by varying the accelerator-to-cathode voltage from 0 to 4500 volts. When the grid is an open slot as shown in the gun cross section of Figure 27-13, the μ of the grid with respect to the accelerator is about 0.2. However, when wires are placed across the grid slot in the appropriate positions, the μ of the grid with respect to the accelerator can be brought up to 5 or more, and this greatly reduces the voltage swing necessary to amplitude modulate the tube.

Whenever the tube is amplitude modulated by varying either the grid or accelerator voltage (while the sole—slow-wave circuit voltage remains fixed) there is a second-order frequency variation called frequency pushing, which is caused by the influence of the bunched team on the slow-wave circuit. When the tube is amplitude modulated by either varying the grid or the accelerator voltages, the pushing is typically 0.6 percent or less of the center frequency as the power varies from full power to zero.

27.3.3 Applications

One of the most important applications of the M-type BWO to the ECM problem is in barrage and spot noise jamming by frequency or amplitude modulation, or both. The noise bandwidth can be varied all the way from a few megacycles to 25 percent of the operating frequency, thus providing a high degree of flexibility in providing watts per megacycle *versus* bandwidth. With the 200-watt series of tubes, one can achieve a noise-power density of about 50 watts/Mc at 4-Mc bandwidth, and this can be spread out to as broad a noise bandwidth as 1800 Mc at X-band. The quality of the noise, in terms of ability to jam sophisticated radars, depends mainly on modulation rates, particularly the frequency-deviation rate. The noise output of an M-type BWO looks like white noise to the most sophisticated radars available today whenever the maximum rate of the noise-frequency deviation exceeds the Dicke-Fix bandwidth. White noise is the most efficient all-around noise that can be used for jamming since it is perfectly random and resembles cosmic noise—the ultimate limitation on sensitivity of all radar receivers. The M-type BWO can also generate many other types of jamming signals for specific applications.

More and more attention is being given to the sole-tuned and modulated M-type BWO. As used in this manner, the main power supply voltage is fixed and the frequency control electrode, the sole, draws very little current, and therefore is a high impedance electrode. The frequency coverage at the present state of the art is 1.2:1, and this is being increased by developments currently under way.

The basic tube efficiency of the sole-tuned tube is essentially the same as that of the magnetron, but since the FM and AM modulation signals are applied to high-impedance electrodes the system efficiencies may be greater than those of the magnetron. The over-all system efficiency in terms of noise watts per pound of equipment may be greater than that of the magnetron, which typically exceeds 1 watt/lb, but the added flexibility of the M-type BWO in providing variable bandwidth barrage or spot jamming, together with deception jamming, make it a very valuable ECM tool.

27.3.4 Future Capabilities

Thus far, the discussion has centered mainly on the 200-watt series of

tubes covering the frequency range of 950 to 11,000 Mc. The feasibility of a similar 400-watt (minimum) series of tubes is demonstrated by the existence of 400-watt tubes, with about 1.4:1 tuning range at L- through S-band and also at X-band. The size and weight of these tubes is no greater at the lower frequencies, but is about 20 percent greater at X-band.

Power outputs of 10 kw have been obtained in M-type BWO's at 200–400 Mc and 10–40 watts has been obtained at about 35 kmc. A development is currently under way for a 50-kw CW tube tunable over a 1.5:1 frequency range at 500 Mc. The attainment of greatly increased power compared to that of existing tubes at any frequency range ultimately demands special means of overcoming delay-line dissipation limits and, at high frequencies, cathode or gun limitations. Generally, the delay-line dissipation problem can be overcome at and below X-band by the introduction of liquid cooling in appropriate delay lines. At K-band and higher, liquid cooling is not feasible because of the small size of delay-line elements, and another method is required. A novel design utilizing "sorting" (Reference 17) has the properties of low delay-line dissipation for a given output power, and offers promise of obtaining greater power than presently possible at K-band and higher. At the present limits of development, cathode or gun limitations are serious only above X-band, and progress is being made on overcoming these limitations by special gun designs. An estimate of objective development capabilities for M-type BWO's is presented in Figure 27-23.

27.4 Backward-Wave Locked-Oscillator or Amplifier (Bitermitron)

The power level available at the present time for crossed-field BWO's is 150 watts, or 400 watts nominal at both S- and X-bands. Increases in this power would, of course, be welcome at any time, and this section describes a tube, the Bitermitron, which is capable of increasing by several-fold the power output now available. At the same time, the Bitermitron has all the advantages of the crossed-field BWO.

The name *Bitermitron* has been given to that class of microwave tubes characterized by the use of two rf terminals, one at each end of the delay line of the non-re-entrant crossed-field backward-wave device. For the applications attempted so far, no attenuation, in addition to that inherent in the delay line itself, has been used. The electron beam, like the delay line, is non-re-entrant.

The Bitermitron is the crossed-field counterpart of the O-type backward-wave amplifier. Up to this time, Bitermitrons have been made for operation at both S- and X-frequency bands by appropriate modifications of a crossed-field BWO. Experience indicates that these tubes can be made for any frequency range for which a basic BWO has been developed. Further
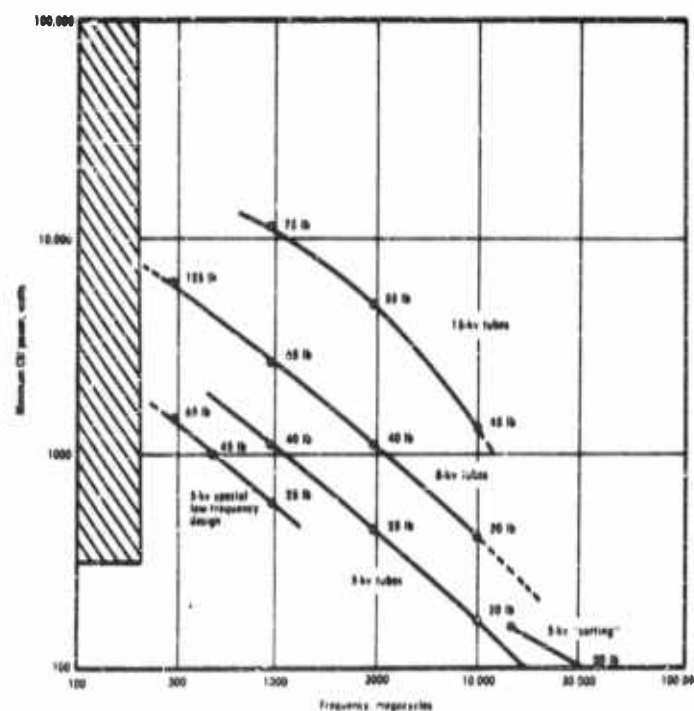
FIGURE 27-23 Objective capabilities of M-type BWO development. Utilization of lumped circuit and liquid-cooled delay-line techniques permit attainment of special combinations of power, size, and weight in cross-hatched region. The notation "lb" is packaged weight.

study of the Bitermitron used as a power booster (Bureau of Ships Contract NObsr 72519) has indicated the type of modifications, particularly of delay-line parameters, needed to improve the tube. With those modifications, the Bitermitron should be a less direct derivative of the prototype crossed-field BWO.

Figure 27-24 is a photograph of an S-band Bitermitron, the QK540, coupled to its BWO driver, the QK625. This arrangement is used for power boosting. The similarity in construction of the two types of tubes is apparent. Their operating voltages and currents are also similar.

The genesis and development of the Bitermitron came from the need for higher output powers than could be obtained conveniently from a single-tube BWO. A second application useful in countermeasures techniques was found when it was noted that, through appropriate circuitry, the Bitermitron

FIGURE 27-24 Bitermitron power-booster system.

could be used both as a voltage-tuned receiver and as a voltage-tuned trans-
mitter. This combined use implies great simplification of equipment.

27.4.1 High-Power Problem

Crossed-field BWO's have received considerable attention because they
can be electronically tuned over a wide (presently 30–50 percent) frequency
band while producing high power with good efficiency (20–40 percent). At
the present state of the art, a 400-watt oscillator can be designed for any
microwave frequency below 11,000 Mc. However, for power outputs above
1000 watts, appreciable difficulties are encountered at frequencies much
greater than 3000 Mc.

Limitations of the crossed-field BWO must be recognized before the merits
of various methods for increasing its power output can be compared. Three
recognized limiting factors in the design of such tubes are: delay-line power
dissipation, cathode current density, and space-charge density.

These three limiting factors are, unfortunately, not independent. For
example, the delay-line dissipation may be increased by reducing the impe-
dance of the delay line, but such a change requires a greater electron-beam
current (increased space-charge density and cathode current density) or an
increase in delay-line length (increased space-charge interaction path).

It is difficult, in the present state of the art, to build a crossed-field BWO
having a conventional interdigital delay line which is capable of producing
1-kw power output over a 30–50 percent frequency band. Without further
basic technological knowledge, advances in existing power-output limits must
probably rely on modification of existing designs. For greater power, one
thinks first of combining the outputs of several existing oscillators in a com-
mon load by connecting the tubes either in parallel or in a series (cascade)
arrangement.

At first glance, parallel operation of the tubes appears impossible because
the power from two or more crossed-field BWO's is not phase coherent and
cannot be successfully combined. However, two methods have been proposed

to overcome these difficulties. In the first method, a small perturbation is formed in the delay lines of $n - 1$ oscillators if n are to be operated in parallel. Such a perturbation, if located at the end of the active delay-line section, has an effect equivalent to that of injecting a locking signal into the slave oscillator from the master oscillator. This latter tube contains no perturbation of its delay line.

A second method of achieving parallel operation of two oscillators involves enclosing the delay lines of both tubes in a single vacuum envelope and passing a single electron beam over the lines. Cross-coupling through adjacency of the delay lines and through the electron beam maintains phase coherence. This parallel delay-line–common-beam arrangement has been used to obtain 1 kw at 3000 Mc. Unfortunately, this technique leads to several practical problems, which include difficulty in constructing the delay lines and matching their impedances. In addition, a rather wide magnet gap is necessary. In contrast to the arrangement involving perturbed delay lines, parallel operation of more than two delay lines in the same vacuum envelope seems impractical.

Series (or cascade) operation of crossed-field backward-wave tubes, the second general method of combining the output of several oscillators, offers several advantages over parallel operation. Among these are: (1) no common electron beam is required so that separate vacuum envelopes may be used; (2) known delay-line manufacturing techniques are applicable; and (3) the magnet gap is no larger than that used in a conventional BWO. The most important advantage of this arrangement, however, is that the output power of a system of cascade tubes can be considerably larger than the sum of the individual oscillator powers. This is true in spite of the losses incurred by transmission of the driver power through the second tube, at least at S-band.

27.4.2 Effect of Insertion Loss

To show the effect of insertion loss on total power output, the curves of Figure 27-25 have been derived for a two-tube driver-Bitermitron system. In each case, the oscillator efficiency has been assumed to be 30 percent and the power booster efficiency to be 40 percent, as defined on the figure.

As the insertion loss becomes higher, the over-all efficiency is reduced because of the decrease in circuit efficiency. In fact, it can be seen that for the 2-db and 3-db curves the over-all efficiency is less than that of either tube. The over-all efficiency for high loss approaches 20 percent, which is just half the Bitermitron efficiency. It is clear that if a number of Bitermitrons, each of finite insertion loss, are connected in cascade to produce very high powers, the rf power-output contribution of the final power booster may just balance its rf loss so that this tube would not contribute to the total power output of the system.
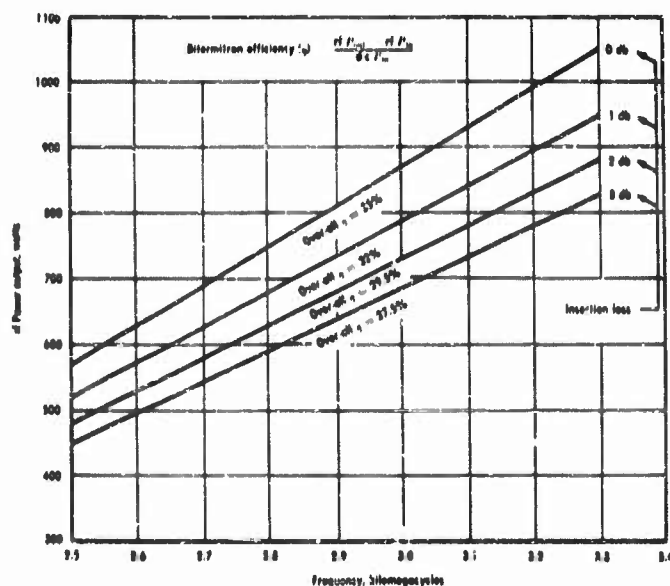
FIGURE 27-25   Effect of insertion loss on system efficiency.

The curves of Figure 27-25 have been drawn for the usual case in which equal d-c powers are applied to the driver and Bitermitron. If this restriction is removed, the above conclusions about the effect of insertion loss may no longer be valid. At 3000 Mc, the insertion loss is quite reasonable, about 1 db for present Bitermitron tubes.

27.4.3 Mode of Operation

With the operating voltages and currents as well as delay-line lengths about the same for both driver and Bitermitron, the power-boosting action of the Bitermitron takes place at currents appreciably greater than its oscillation starting current. For this reason, it is accurate to say that the Bitermitron, in this application, is a locked oscillator rather than an amplifier. The free-running frequencies, voltages, power outputs, etc., of the Bitermitron are essentially those of its prototype BWO.

Locking of the Bitermitron free-running frequency to that of the drive frequency will occur at an input power dependent upon the separation of the frequencies of the driver and the free-running Bitermitron. When the rf drive power to the Bitermitron is increased, the Bitermitron power output at the driver frequency increases until, with full drive power, the system power output may be 3 to 4 times that of either tube alone.

Since either tube will oscillate by itself, the use of a driver-Bitermitron power-booster system offers some assurance of continuity of operation in any critical application. The power outputs obtainable from either one of the tubes would be one-third to one-fifth the normal output of the two-tube system.

It should be noted that the above discussion implies a BWO driver and a Bitermitron of similar construction except for those changes required in the Bitermitron. In this case, the voltage-tuning feature is preserved since it has been shown experimentally that the locking range or usable bandwidth is sufficient to permit use of a common anode supply. The Bitermitron power booster is, however, entirely suited for enhancement of power produced by other types of drivers, such as a magnetron. The magnetron or other driver may be of fixed frequency, or may have a mechanically adjusted tuning range within that of the Bitermitron.

27.4.4 Medium-Power S-Band System

The early development of the S-band Bitermitron was accomplished through modification of a basic 150-watt crossed-field BWO, the QK481. This oscillator has since undergone production refinement and is now designated the QK625. The refinement work brought about detailed improvements in the mechanical design of the QK481 and permitted an increase in its minimum power output to 180 watts.

Development work on the associated Bitermitron, the QK540, has kept pace with the refinement program resulting in the QK625. The combination of the QK625 driver with the QK540 Bitermitron forms a basic, medium-power, S-band system. The two tubes of this arrangement are similar in size, shape, weight, and electrical characteristics. Pictured in Figure 27-24 is the QK625 BWO coupled to drive a QK540 Bitermitron power booster.

The similarity of the QK540 and the QK625 is emphasized by Figure 27-26, which shows the power output and efficiency of a QK540 being operated as a free-running oscillator. This performance approximates that which would be obtained with typical QK625 BWO's. When the QK540 is operated with small amounts of injected power, the characteristic curves can be expected to appear as in Figure 27-27. The locking characteristics and 3-db bandwidths are shown for two values of injected power; and it is apparent that both the QK540 locked-oscillator efficiency and its bandwidth become greater with increasing drive power. The power-output curves here are reasonably symmetrical about the power maximum, a condition that does not hold true for higher drive powers.

Figure 27-28 illustrates the behavior of several Bitermitron characteristics as a function of drive power. These data are for a midband frequency of
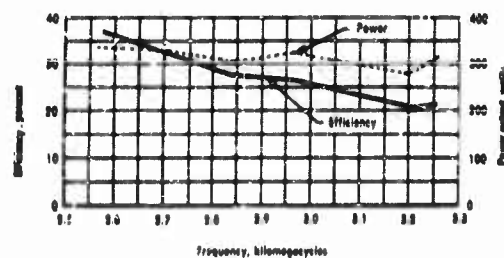
FIGURE 27-26 QK540 Bitermitron used as free-running oscillator (with external attenuator).
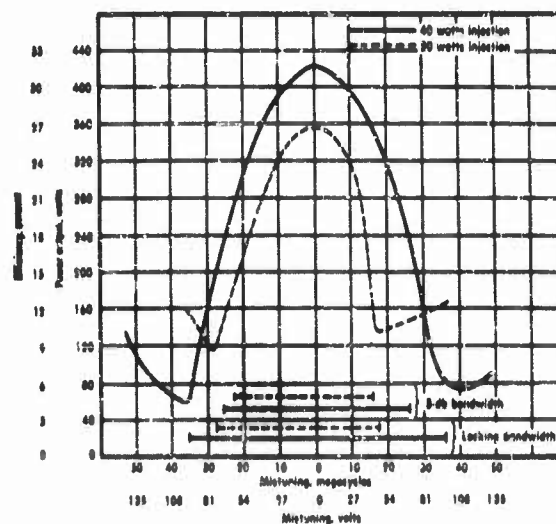
FIGURE 27-27 Typical injection-locked-oscillator performance of S-band Bitermitron (QK540).
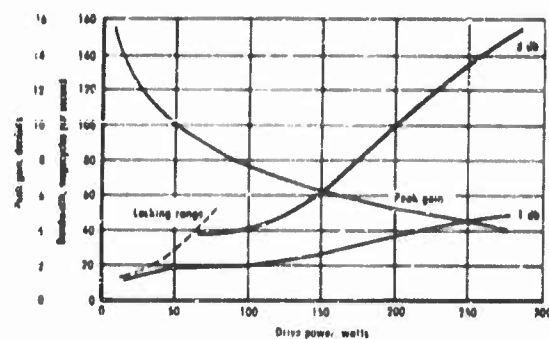
FIGURE 27-28 Bandwidth and gain *versus* drive power of S-band Bitermitron (QK540).

27-34

2900 Mc. The locking-range curve has been derived for only a small range of drive powers since it is believed that most users will be interested only in the 1-db or 3-db bandwidths. Experiment shows that over nearly the entire locking range outside the 3-db bandwidth the output power is no greater than that which can be obtained from a single tube. The smallest drive power attempted in these experiments, about 10 watts, yielded a peak gain of approximately 16 db. In this region of drive power, the system becomes difficult to operate unless the driver is insensitive, or is made insensitive, to reflections. Such insensitivity is necessary because an appreciably mismatched load may cause the reflected power to be greater than that injected.

QK625 and QK540 tubes can be obtained separately and may be coupled to form a two-tube driver-Bitermitron system. However, these tube types are also available in what is called a "matched-pair" system. This consists of a QK540 Bitermitron individually adjusted to track with a particular QK625 driver. The tracking is secured by factory adjustment of the QK540 magnetic field strength while keeping common anode and sole voltages on both driver and booster. A set of operating characteristics detailing the system performance is supplied with each matched-pair system. Unrelated driver and booster tubes can be made to track over the band with a common anode supply, but a spread of sole voltages between the tubes may be necessary. Work is progressing to standardize the magnetic field strength of the QK625 and the QK540 tubes in order to eliminate the need for matched pairs.

Operating characteristics for a matched-pair system are illustrated in Figure 27-29. Considering that the driver power may have been about 225 watts at the low-frequency end of the band and increased to perhaps 300 watts at the high end, the corresponding booster contribution to the power
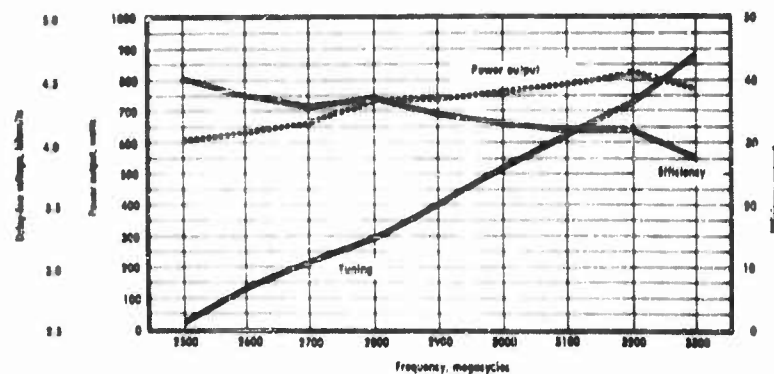
FIGURE 27-29 Operating characteristics of S-band driver-Bitermitron system (QK625/QK540).
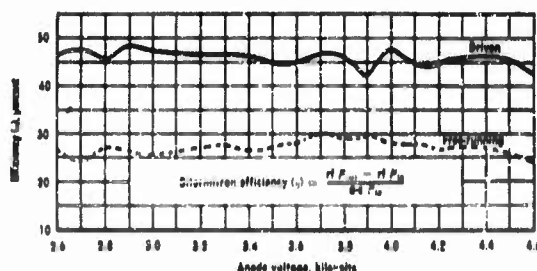
FIGURE 27-30 Free-running and driver efficiencies of an
S-band Bitermitron (QK540).

output varied between 375 and 500 watts. The rated minimum output power
of a QK625/QK540 medium power S-band system is 450 watts. Figure 27-30
shows the increase in efficiency of a typical QK540 Bitermitron as drive
power is applied.

Some work has been done at S-band with a system comprising a driver
followed by two Bitermitron power boosters. However, the insertion loss of
the QK540 tubes begins to reduce the over-all system efficiency although
this loss is compensated in part by the increases of efficiency of the second
power booster. It seems clear that the second Bitermitron in such a chain
requires different design considerations to make the most effective use of the
large drive powers available at its input.

For the above reasons, as well as for reasons of system complexity, the
use of three tube arrangements does not appear practical at present.

27.4.5 Higher-Power and Higher-Frequency Bitermitrons

Further work (Bureau of Ships Contract NObsr 72519) on the Bitermitron
has resulted in a higher-power S-band Bitermitron system, the QK625/
QK618, and an X-band Bitermitron system, the QK634/QK605.

Although the QK625/QK618 system is specified to produce a minimum of
1 kw, generally 1200 to 1700 watts is obtained over the frequency range of
2500 to 3300 Mc. In this system, both the driver and the Bitermitron are
programmed in beam current as a function of anode voltage.

Bitermitrons at the 450-watt level have been also developed at X-band and
at other S-band frequency ranges.

27.4.6 The Superregenerative Bitermitron

The high-power features of the Bitermitron which are particularly suited
to countermeasures equipment can be used most effectively in a jammer if
frequency acquisition of the victim radar can be simplified.

With the Bitermitron, high-power amplification of microwave signals is

possible, but low-power amplification has been tried as well. It is not possible to lock the oscillation of the Bitermitron with signals having a power level of 1 milliwatt or less. However, if the anode current is kept at a value lower than that required to develop backward-wave oscillation, regenerative voltage-tunable amplification can be achieved. This regenerative gain follows fairly closely the usual laws of regenerative amplifiers in that gain can be exchanged for bandwidth, and vice versa.

Since regenerative gains are not particularly useful because of extreme operational instability at high gains, superregenerative techniques were tried.

The superregenerative amplifier has a gain-bandwidth product which is much greater than that of the regenerative amplifier. When each operational phase of superregeneration is considered, the reason for this large improvement becomes more evident.

In a superregenerative amplifier, signal energy and oscillatory energy are fed to a single tuned circuit. The ideal quenching cycle of such an amplifier consists of forcing the amplifier into a quenched state for a short period and then placing it in a regenerative, or "sensitive" state for a much longer period. Following this, the amplifier is driven into full oscillation. The duration of oscillation buildup is governed by the received signal amplitude in the tuned circuit. Then the complete quench cycle is repeated. The period of this cycle should be much less than that of the highest modulation frequency.

The bandwidth of the amplifier will be determined by the rate at which the quench voltage travels in and out of regenerative amplification and oscillation. In general, the faster the rate, the wider the bandwidth, because for the same quench waveform, less time is spent in the regenerative or sensitive part of the cycle at the higher quench frequencies.

In an experiment on the Bitermitron as a superregenerative amplifier, the quench voltage was applied to the control grid. The signal chosen to be amplified was a 10-μsec pulse applied at a 1-kc repetition rate. This signal allowed 10 samples per pulse to be taken by using a 1-Mc sine-wave quench frequency. The usual multiple resonances associated with superregenerative amplifiers were not apparent in the Bitermitron.

The range in anode current over which good amplification was obtained was quite small; consequently, slight variations in current caused large variations in gain. The use of the superregenerative Bitermitron in a countermeasure system, where band sweeping would be used for radar frequency acquisition, requires the receiver to have constant sensitivity over the band. Such constant sensitivity is not obtained when the Bitermitron is superregenerative because the oscillation starting current, and therefore the current for equal regenerative gain is not constant over the band. Therefore, means

for establishing an effective automatic sensitivity control was necessary for good frequency-acquisition characteristics as well as for reliable measurements of gain, bandwidth, and sensitivity. Automatic sensitivity control was obtained by amplifying the quench signal component from the crystal detector at the Bitermitron output and rectifying it; the resultant d-c voltage was fed back to the Bitermitron grid while mainaining suitable isolation between the d-c voltage and the quench signal voltage on the grid. The Bitermitron grid has a transconductance of approximately 200 μa/volt, which is sufficient to maintain reasonably constant sensitivity over the band when using a maximum rectified quench voltage of 100 volts d-c. This arrangement is also an effective automatic gain control.

The signal-to-noise ratio at the Bitermitron output was optimized and the quench signal component was removed from the signal by amplifying the signal in a multistage amplifier of bandwidth $1/T$, where T is the pulse width of the signal. The amplifier bandwidth was about 100 kc. Experimental results obtained with a laboratory model of a receiver built according to these principles are given below.

Input-signal powers from -35 dbm near the low end of the band (2600 Mc) to -51 dbm at the high end (3300 Mc) were measured in one superregenerative Bitermitron for an output signal barely discernable in noise as viewed on an oscilloscope. Power gains of this tube varied from 50 db to 71 db at the low- and high-frequency ends of the band, respectively. These sensitivities are approximately the same as those of crystal video receivers currently used in countermeasures systems. Some improvements in these sensitivities and gains could be realized if the quench waveform were tailored for better signal to noise ratios.

For the purposes of illustrating the use of this unique tube, the superregenerative Bitermitron, a block diagram of a possible countermeasures system is given in Figure 27-31. Some of its advantages over a more conventional system are the elimination of many extra components needed in the frequency-acquisition portion of the jammer. These components include the mixer, the i-f strip, the local oscillator, and their power supplies, as well as the mixer and discriminator for automatic frequency control. Alignment error of the transmitter frequency with respect to the received frequency is eliminated because, at any power level, the Bitermitron radiates at substantially the center frequency of the passband of the superregenerative amplifier.

Operation of the circuit shown in Figure 27-31 proceeds in the following manner.

In the receive condition, the quench oscillator signal is applied to the Bitermitron grid. A high level of sensitivity is maintained by means of the automatic sensitivity-control circuits while the tube is swept over the band.
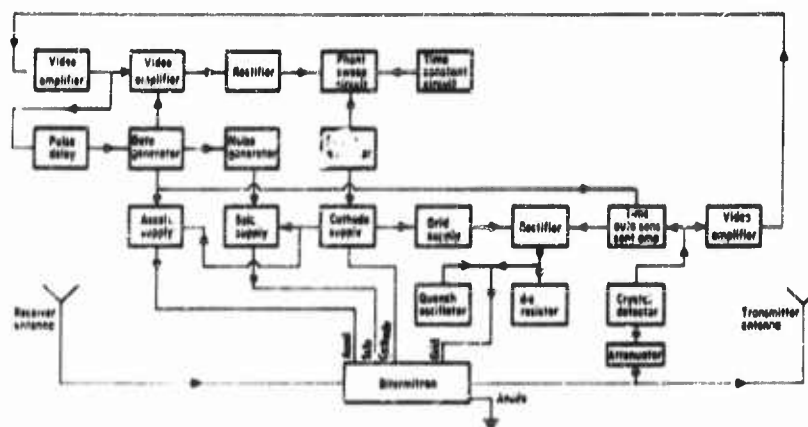
FIGURE 27-31 Possible countermeasures system using the Bitermitron.

When a radar signal within the band appears at the receiver antenna, it is amplified by the Bitermitron and detected by the crystal. From here, the resulting video signal is passed through a video amplifier, whose high-frequency cut-off is designed for optimum signal-to-noise ratio and to remove the quench-signal component normally present at the crystal output. After amplification, the video signal is rectified and filtered to develop a d-c voltage bias which is applied to the grid of a Phantastron sweep circuit.

The Phantastron, with no bias applied to its grid, normally sweeps the anode supply which, in turn, sweeps the Bitermitron over its band. As the Bitermitron is swept, its bandpass moves to include the frequency of the victim radar, and the negative bias on the Phantastron grid increases with the increasing video signal. The increasing negative bias reduces the sweep rate and eventually stops the sweep when the radar signal frequency is on the skirt of the Bitermitron selectivity curve. The phase relationships are such that, if the Bitermitron frequency tries to increase, the resulting decrease in video signal will reduce the bias on the Phantastron grid to decrease the frequency. If the Bitermitron frequency tries to decrease, the resulting increase in video signal will increase the bias on the Phantastron grid to increase the frequency.

The video output signal is also used to trigger a gate generator, but is delayed sufficiently to allow the automatic frequency-control action to take place first. The gate generator increases the accelerator anode voltage of the Bitermitron until sufficient anode current is drawn to allow the tube to go into full power oscillation. Simultaneously, the gate generator turns on the noise-source signal, which is amplified and applied to the sole electrode.

While the Bitermitron is in the transmit condition, no received signal is available to maintain the tube locked on the victim frequency. Therefore, a time-constant circuit is required to maintain the Phantastron plate voltage at the value that was established during the automatic frequency-control action. The delay of the gate generator trigger is sufficient to allow this time-constant circuit to charge fully to the correct value before jamming cycle is initiated. The gate generator must also disable the automatic sensitivity-control circuit during jamming so that the crystal output will not cause any changes in Bitermitron grid voltage which would vary the jamming power.

After a jamming cycle, the duration of which is determined by the gate-generator pulse duration, the Bitermitron is returned to the receive condition, and sweeping resumes from the frequency where the jamming had taken place.

27.5 The Voltage Tunable Magnetron

27.5.1 Description of Tubes

The VTM (voltage-tunable magnetron oscillator) consists of three basic components: (1) the tube subassembly (an example is shown in Figure 27-32); (2) the rf circuit or cavity; and (3) the magnet. These three com-
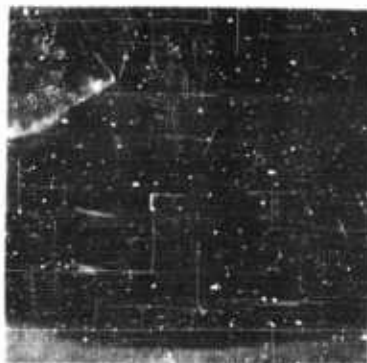
FIGURE 27-32 Voltage-tunable magnetron tube subassembly.
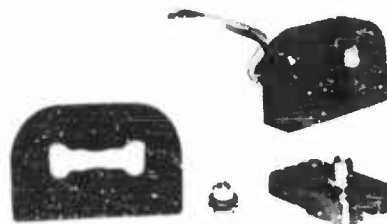
FIGURE 27-33 Voltage-tunable magnetron components.

ponents are aligned and locked together to form an integral assembly with an rf output connector and d-c input leads. Typical components and assembly are shown in Figure 27-33. Design differences made in any of the three basic components permit a wide variation to be made in the operating characteristics such as frequency range, electronic tuning range, power output, efficiency, and noise.
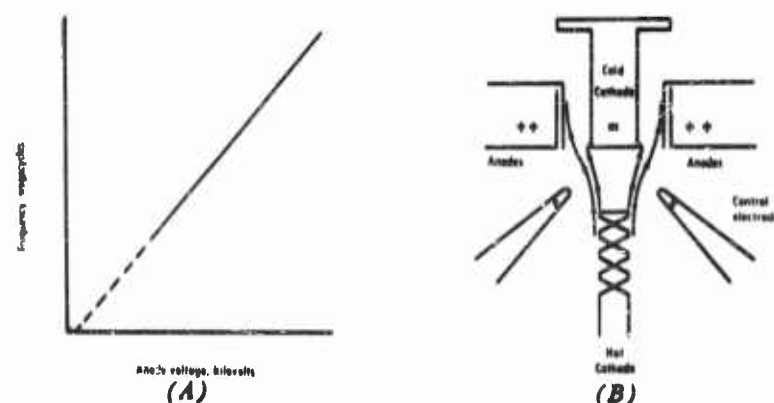
FIGURE 27-34 Voltage-tunable magnetron tuning characteristic and tube subassembly schematic.

A schematic of a VTM tube subassembly is shown in Figure 27-34b. The anode structure illustrated is interdigital and the wave-beam interaction takes place in the cylindrical interaction space formed by this anode and a nonemitting cathode or sole. A whirling tubular sheath of electrons is obtained from the hot cathode and controlled and injected into the interaction space by means of the injection (or control) electrode. The hot cathode illustrated is made of thoriated tungsten and it is protected from back heating effects by the injection electrode. It may therefore have space-charge-limited operation and can be designed for very long life. The injection electrode collects very little electron current and may be used to amplitude modulate or pulse the rf power output as well as to control the injected beam current. The anode-sole voltage determines the frequency of oscillation, and the voltage-frequency relation, as illustrated in Figure 27-34a, is proportional and linear.

Two wideband VTM's which make use of standard E-magnet construction are shown in Figure 27-35. On the left is the GL-7398, an S-band VTM that produces 2 watts minimum power from 2200 to 3850 Mc. On the right is an L-band VTM that produces approximately 1 watt minimum from 1000 to 2300 Mc. These units weigh $3\frac{1}{2}$ and 3 lb respectively.

The tube subassembly itself is quite small; a penny can cover it in any view. It is constructed of ceramic and metal parts that are joined by high-temperature seals. This construction results in an assembly that is extremely rugged. The rf circuit or cavity in which the tube subassembly is mounted can vary considerably in form. Its function is to provide for the electronic interaction an impedance that will produce the desired characteristics from the assembly. The VTM is noteworthy among electronically tunable microwave tubes in that a great variety of electrical performance can be obtained
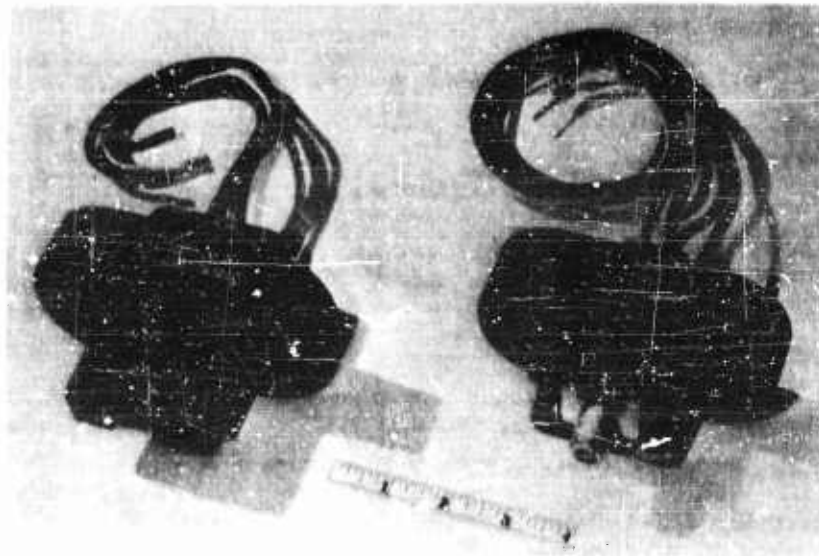
FIGURE 27-35   Voltage-tunable magnetron tubes for wideband operation through S and
L-Band.

by operating the same tube subassembly with external circuits having dif-
ferent impedance characteristics. This versatility will be demonstrated later.
The VTM has been packaged with various types of magnets. These include

FIGURE   27-36   Bowl-magnet
voltage-tunable magnetron.
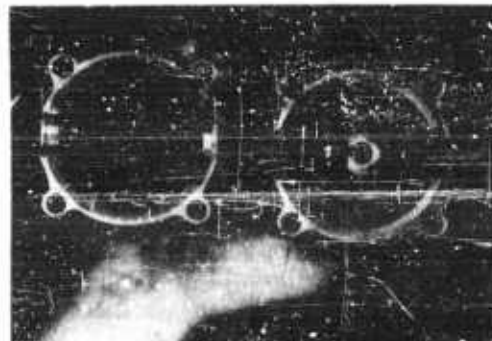
FIGURE 27-37   Internal construction of bowl-magnet
voltage-tunable magnetron.

E-magnets, as already shown, C-magnets, and bowl-type magnets. A package utilizing bowl-magnet construction is shown in Figure 27-36. The neat, compact, inner construction is shown in Figure 27-37. This unit weighs 1.3 ib and delivers 5 watts of power over 250-Mc band in the S-band frequency range. This unit, like others in development, operates without external cooling and is capable of withstanding severe missile shock, vibration, high altitude, and other environmental conditions.

27.5.2 Operating Characteristics

Typical performance of a broad S-band production model GL-7398 is shown in Figure 27-38. Typical performance for a broad L-band VTM is shown in Figure 27-39. It should be noted that the tuning characteristics of
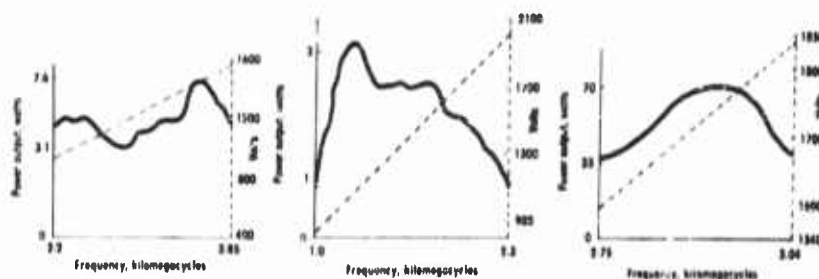
FIGURE 27-38 Power output and voltage-tuning characteristics for wideband S-band voltage-tunable magnetron GL7398.

FIGURE 27-39 Power output and voltage-tuning characteristics for wideband L-band voltage-tunable magnetron.

FIGURE 27-40 High-power, high-efficiency narrow-band operation of a voltage tunable magnetron.

these tubes are linear and that the frequency is in direct proportion to the applied anode voltage. This means that a 2:1 change in voltage produces a 2:1 change in frequency. The "average" efficiency over the band for both described VTM's is approximately 25 percent. Wideband operation is obtained by using a circuit that has an extremely low Q impedance characteristic. With a circuit that has a higher Q impedance characteristic, higher powers can be obtained over smaller electronic tuning ranges. This "narrow" band operation may be obtained with a center frequency anywhere in the operating range of the tube subassemblies. Typical data for narrowband operation is shown in Figure 27-40. Approximately 40 watts minimum is obtained over a 10 percent electronic tuning range at an efficiency in excess of 50 percent. A tube subassembly identical in appearance with the GL-7398 tube subassembly is used for this narrowband, high-power operation.

In addition to complete VTM coverage on the L- and S-band ranges, narrowband VTM's having lower output powers have been made for operation

in the C-band region. The present frequency limit for voltage-tunable magnetron power at the 1-watt level is approximately 6000 Mc. Voltage-tunable-magnetron power at the milliwatt level in the X- and K-band regions may be obtained by making use of S- or L-band tubes in conjunction with harmonic generators. Sufficient power is obtained for local oscillator applications, and the extremely linear voltage-frequency characteristic is preserved. A number of these VTM harmonic generator combinations are in use.

The VTM is capable of extremely rapid frequency modulation since the tube subassembly and circuit are short compared to the wavelength and since the velocity of all portions of the beam may be changed simultaneously. It may be frequency modulated at frequencies approaching its oscillation frequency. Other practical considerations, in addition to the linear tuning characteristic, include the low input capacity and large modulation sensitivity which considerably lessen the demand on the modulator that must be used when high and varying frequency rates are desired. Relatively small voltages at low input are needed to frequency modulate the VTM.

The amplitude of the output signal of the VTM is controlled by the voltage on the control electrode. The tube may be pulsed on and off at extremely high rates by pulsing this voltage. The control electrode collects essentially no electron current. Present data, though limited, indicates definite amplitude-modulation capabilities. In particular, a VTM with an electronic tuning range of 10 percent (i.e., a narrowband version) has been varied in amplitude over a power range of better than 15:1; this variation was essentially linear with the control electrode voltage.

27.5.3 Applications

The VTM is well suited for at least three important roles in the countermeasures field. These are (1) local oscillators; (2) drivers; and (3) output tubes. As local oscillators for swept receivers, the linear and proportional tuning characteristic makes possible features that cannot easily be obtained with other forms of voltage-tuned tubes. For such applications one tube can suffice for coverage of many frequency bands by making use of harmonic generators and power dividers. The VTM is relatively free of spurious signal outputs inasmuch as the rf structure is electrically too short to support other transmission modes in the operating frequency bands. There is also sufficient data on noise to indicate that the units currently available are satisfactory for local-oscillator operation without further special design, and they have already been used successfully in such applications.

The VTM may be used to drive a wideband amplifier for either spot or expandable barrage jamming. Because this type of tube has the rather high power output of 1-2 watts over an octave frequency range, the VTM can be

used to drive a traveling-wave-tube power amplifier and control hundreds of kilowatts of power over a 2:1 frequency band. It can be amplitude or frequency modulated, or both, and with the advantages of easy electronic tuning at low power a high degree of system sophistication becomes practical.

Finally, the narrowband version of the VTM can produce 50 to 100 watts of output power at efficiencies approaching 60 percent and with 10 to 15 percent electronic tuning. It is well adapted for direct output use in countermeasures systems requiring versatility, ruggedness, and light weight, and in particular appears especially attractive for drone and missile countermeasures systems.

27.5.4 Future Capabilities

The capabilities of the VTM may be extended in a variety of directions. First, the existing broadband S- and L-band VTM's can be supplemented by similar octave-tuning VTM's covering the entire 100-mc to 6000-mc region. In addition, a different type of tube subassembly should permit extension of the narrow-band versions of 10 to 15 percent bandwidth into x-band at a level of several watts. The efficiency for tubes of this bandwidth may be expected to advance into the region of 60 percent, permitting air-cooled units of 100-watt output power. At the lower frequencies, larger tuning ranges at even higher powers may be expected. The VTM noise figure is now comparable with that of any other electronically tuned device. Further reduction can be expected; to date little effort has been expended toward this end.

REFERENCES

1. Collins, George G., ed., *Microwave Magnetrons*, M.I.T. Radiation Laboratory Series, McGraw-Hill, New York, (1948).

2. Fisk, J. B., H. D. Hagstrum, and P. L. Hartman, "The Magnetron as a Generator of Centimeter Waves," *Bell System Tech. J.*, Vol. 25, p. 167 (1945).

3. Slater, John C., *Microwave Electronics*, Van Nostrand, New York, (1950).

4. Wilbur, D. A., *et al.*, *C-W Magnetron Research*, Final Report, Contract No. W-36-03vsc-32279, Research Laboratory, General Electric Company (April 1, 1950).

5 Needle, J. S., *The Insertion Magnetron* Project M 921, The University of Michigan, Dept. of Electrical Engineering, Technical Report No. 11 (August 1951).

6. Welch, H. W., Jr., "Prediction of Traveling-Wave Magnetron Frequency Characteristics," *Proc. I.R.E.*, Vol. 41, No. 11 (November 1953).

7. Boyd, J. A., "The Mitron—An Interdigital Voltage-Tunable Magnetron," *Proc. I.R.E.*, Vol. 43, No. 3 (March 1955).

8. Kompfner, R., "The Traveling-Wave Tube," *Wireless Eng.*, Vol. 24, pp. 255-266 (1947).

9. Pierce, J. R., and L. M. Field, "Traveling-Wave Tubes," *Proc. I.R.E.*, Vol. 35, pp. 108-111 (1947).

10 Warnecke, R. R., *et al.*, "The Magnetron-Type Traveling-Wave Amplifier," *Proc. I.R.E.*, Vol. 38, No. 5 (May 1950).

11. Warnecke R. R., *et al.*, "The M-Type Carcinotron Tube," *Proc. I.R.E.*, Vol. 43, No. 4 (April 1955).

12. Dench, E. C., "A Voltage-Tuned High-Power Microwave Oscillator," *Proc. Natl. Electronics Conf.* (1954).

13. Hull, J. F., "The L-3147 Backward-Wave Oscillator," *Trans. 1957 Defense Department Symposium on Electronic Countermeasures*, The University of Michigan, Ann Arbor, Mich., pp. 110-127.

14. Hull, J. F., and P. W. Crapuchettes, "Some Recent Advances in ECM Microwave Tubes," *Trans. 1959 Defense Department Symposium on Electronic Countermeasures*, The University of Michigan, Ann Arbor, Mich.

15. Currie, M. R. and J. R. Whinnery, "The Cascade Backward-Wave Amplifier," *Proc. I.R.E.*, Vol. 43, p. 1617 (1955).

16. Kluver, J. W., "Aspects of M-type Interaction with Particular Reference to the Backward-Wave Magnetron Amplifier," presented at the International Convention on Microwave Valves, May 21, 1958, London; published in the *J. Inst. Elec. Eng.* (London), p. 605-608 (November 1958).

17. Osepchuk, J. M., "Sorting in M-type Traveling-Wave Tubes," presented at the Annual Electron Tube Conference, Mexico City, June 1959. To be published.

28

Ferrimagnetic, Gaseous Electronic, and Ferroelectric Devices

H. W. WELCH, Jr., A. L. ADEN

This chapter treats separately ferrimagnetic, gaseous electronic, and fer-roelectric ECM devices for use below microwave frequencies, specifically below 1000 mc, and at microwave frequencies. For completeness, a brief review of the basic theory of ferrites is given in connection with the micro-wave applications since the microwave properties of these materials are not familiar to many engineers. Since the low-frequency applications are based on the well-known nonlinear properties of ferromagnetic material and the similar nonlinearity in ferroelectric materials, treatment of the basic theory is not considered necessary.

I. Applications of Ferrite and Ferroelectric Materials below 1000 Megacycles

28.1 Nonlinear, Low-Loss, and Temperature Characteristics of Fer-rite and Ferroelectric Ceramics

The basic characteristics which make ferrite ceramics useful for application in electronic countermeasures equipment are the following:

(A) The nonlinear relationship between magnetic permeability μ and applied magnetic field H characteristic of ferromagnetic materials.

(B) The low loss of these materials, compared to that of ferromagnetic metals, which makes possible their application at higher frequencies.

The characteristics of both ferromagnetic and ferroelectric materials are subject to variation with temperature which, if uncontrolled, limits their application. The hysteresis effect, described below, further limits application in specific instances.

These general characteristics are illustrated and useful parameters are defined in Figures 28-1 through 28-6.

A typical nonlinear characteristic showing hysteresis and defining incremental permeability and dielectric constant is given in Figure 28-1. Figure 28-2 shows the typical variation of μ_Δ or ϵ_Δ with biasing field H or E. In the mks practical system, the units of the various quantities shown are as follows:

Biasing field	H	amperes/meter	E	volts/meter
Flux density	B	webers/meter^2	D	coulombs/meter^2
Incremental permeability or dielectric constant	μ_Δ	henries/meter	ϵ_Δ	farads/meter

Typically μ_Δ and ϵ_Δ are referred to as *relative* values with the free-space values as a reference. Thus

$$\mu_\Delta \text{ relative} = \frac{\mu_\Delta}{\mu_0}$$

where $\mu_0 = 4\pi \cdot 10^{-7}$ henries/meter

$$\epsilon_\Delta \text{ relative} = \frac{\epsilon_\Delta}{\epsilon_0}$$

where $\epsilon_0 = \frac{1}{36\pi} \cdot 10^{-9}$ farads/meter.

B is frequently given in gauss and H in oersteds where 1 gauss $= 10^{-4}$ webers/square meter and 1 oersted $= 10^3/4\pi$ ampere turns/meter.

Unfortunately a variety of subscripts and units are in common usage. The reader is warned to determine which is being used when referring to tabulated information, data, or other references.

The terms μ_0 and ϵ_0 are frequently used to refer to initial values as $\triangle H$ or $\triangle E$ and B or D go to zero.

The losses in magnetic and dielectric materials are accounted for by giving μ and ϵ a complex value. Thus

$$\mu^* = \mu' - j\mu'' \qquad \text{and} \qquad \epsilon^* = \epsilon' - j\epsilon''$$

μ'/μ'' and ϵ'/ϵ'' are intrinsic Q of the material; $1/Q$ is the loss tangent.
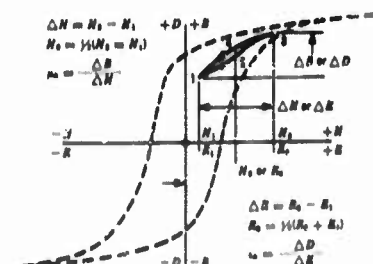
FIGURE 28-1 Definitions of magnetic and dielectric parameters.
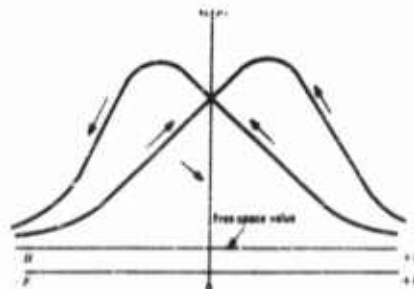
FIGURE 28-2 Variations of μ_Δ and ϵ_Λ with biasing field.

FIGURE 28-3 Relative permeability and relative constant as a function of frequency for magnesium-manganese ferrite.
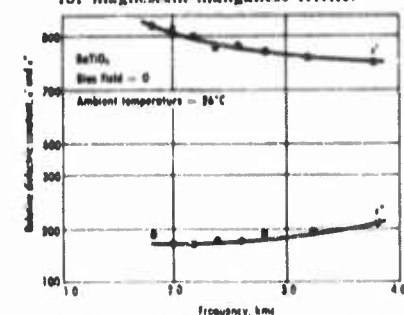
FIGURE 28-4 Real part of permeability of ferrite "O-3" as a function of temperature (imaginary part of data not immediately available).
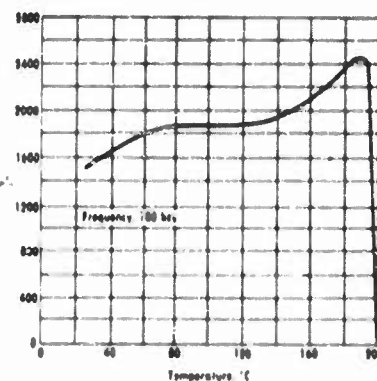
FIGURE 28-5 Relative dielectric constant vs frequency, typical ferroelectric material

FIGURE 28-6 Relative dielectric constant vs temperature, typical ferroelectric material.
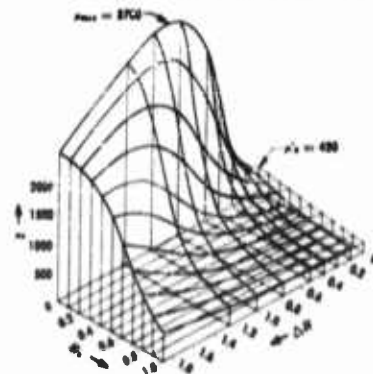
FIGURE 28-7   Surfaces of μ_Δ, H_0, and ΔH for ferramic H ferrite material (Fig. 7, Tech Rept. 37, EDG, The University of Michigan, Ann Arbor).
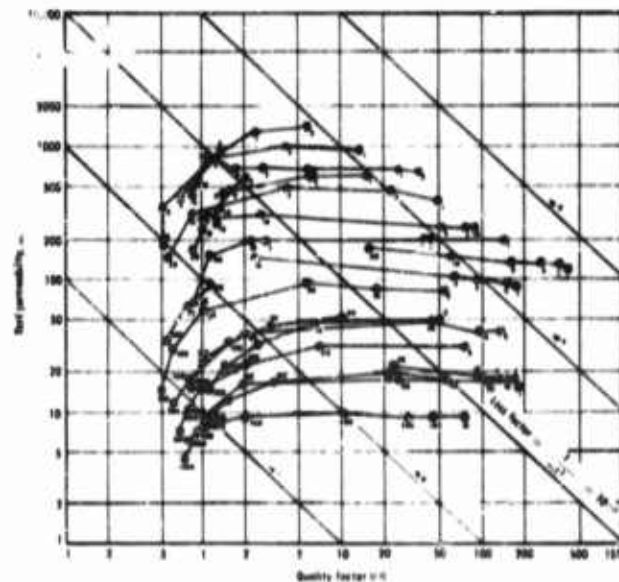
FIGURE 28-8 μ-Q plot of several ferrites.

CORE TYPE (Experimental Material)

1	Ferroxcube B-5	11	EDG A-231-5
2	EDG D-150-1	12	EDG A-290-4
3	EDG D-143	13	GC Q
4	EDG C-80-1	14	EDG E-101-1
5	Ferroxcube B-4	15	EDG D-121-1
6	Ferroxcube B-3	16	GC F-146-D
7	Ferroxcube B-2	17	GC N
8	Ferroxcube B-1	18	GC F-174-E
9	EDG D-142	19	GC F-141-E
10	EDG A-105	20	GC F-34-A
		21	Phillips $CO_2Ba_2FE_{24}O_{41}$

Frequency (mc) shown for each measurement.

FIGURE . efinitions of magnetic and dielectric parameters.

FIGURE 28-2 Variations of μ_Δ and ϵ_Δ with biasing field.

FIGURE 28-3 Relative permeability and relative constant as a function of frequency for magnesium-manganese ferrite.

FIGURE 28-4 Real part of permeability of ferramic "0-3" as a function of temperature (imaginary part of data not immediately available).
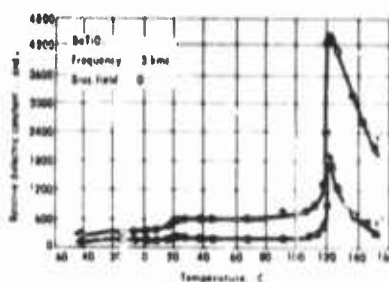
FIGURE 28-5 Relative dielectric constant vs frequency, typical ferroelectric material.

FIGURE 28-6 Relative dielectric constant vs temperature, typical ferroelectric material.
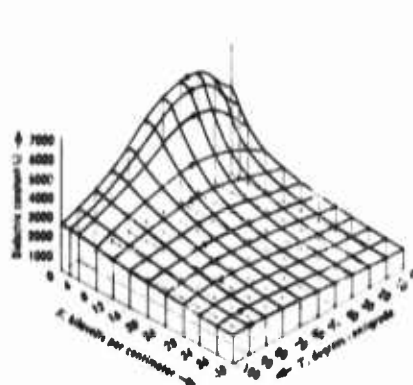
FIGURE 28-9 ε-T-E surface for Aerovox "Hi-Q" 40 (Fig. 1, Tech. Rept. 53, RDG, The University of Michigan, Ann Arbor). $E_{a-c} = 0.005$ kv-cm⁻¹ (rms) at 1 kc. Cycling field, 1 polarity. Cycling field rate = 100 kv-cm⁻¹ -min⁻¹. Data plotted only for E_{d-c} increasing.
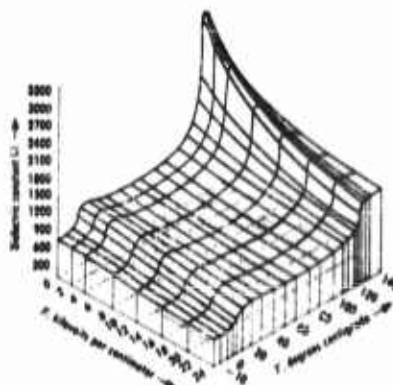
FIGURE 28-10 ε-T-E surface for Centralab D-13 (Fig. 3, Tech. Rept. 53, RDG, The University of Michigan, Ann Arbor). Centralab batch No. H51-33, #151-958. E_{a-c} = 0.1 kv-cm⁻¹ (rms) at 1 kc. Cycling field, 1 polarity. Cycling field rate, 70 kv-cm⁻¹ min⁻¹. Data plotted only for E_{d-c} increasing.
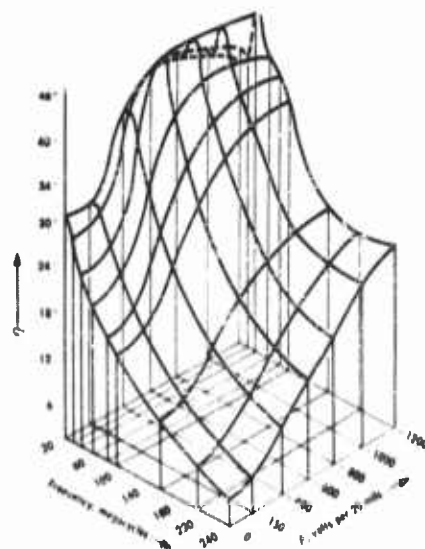
FIGURE 28-11 Q-E-F surface (Reference Section 3, p. 841, Proceedings of the National Electronics Conference, 1955, Vol. XI).

Typical variation of relative μ^* and ϵ^* with frequency and temperature is given in Figures 28-3 through 28-6. Temperature coefficients of less than 1% relative change in μ or ϵ per degree centigrade have been obtained in commercially available materials. This still makes necessary the use of oven temperature control in many applications.

Ferrites are ceramics composed of three or more oxides one of which is iron in a spinel crystalline structure. Ferroelectric ceramics are usually titanates, niobates or zirconates of barium, strontium, cadmium, or lead with additives to provide particular properties. Literally thousands of compositions of ferrite and ferroelectric materials are possible and hundreds exist. This makes the matter of selection for particular applications difficult. Figures 28-7 through 28-11 give data on typical materials in a form that is a convenient aid to the selection process. Data in this form are usually not available from the manufacturer in the completeness shown here so it is up to the user to make additional measurements. Figure 28-8 shows a special form of data presentation that is useful in selection of ferrite materials. Similar data would be helpful in the selection of ferroelectric materials but is not readily available.

28.2 Circuits Using Ferrite and Ferroelectric Ceramics

28.2.1 Variation in Inductance and Capacitance With Biasing Field

At frequencies below 1000 mc the possibility of practical application of ferrite and ferroelectric materials most pertinent to electronic countermeasures is made evident by examination of Figures 28-12 and 28-13. Figure 28-12 shows the variation in impedance of a ferrite-cored inductance with a
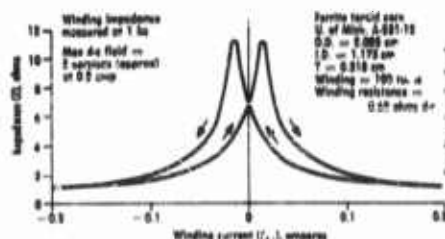
FIGURE 28-12 Impedance in ohms vs bias current for ferrite core toroid, parallel fields (Fig. 14, Tech. Rept. 61, EDG, The University of Michigan, Ann Arbor).
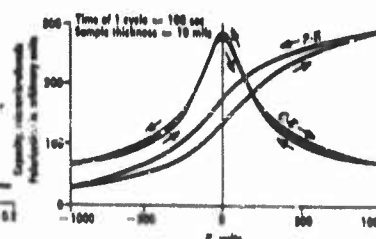
FIGURE 28-13 Capacity in micromicrofarads vs biasing voltage for ferroelectric capacitor (Fig. 10, Tech. Rept. 61, EDG, The University of Michigan, Ann Arbor).

d-c biasing current in the winding. Figure 28-13 shows the variation in capacitance of a ferroelectric capacitor with applied d-c voltage bias (compare

to Figure 28-2). This variation in inductance or capacitance, subject to the limitations of hysteresis effects, temperature effects, and variation with frequency described above, can be utilized in electronic tuning of oscillators, mixers, and rf stages, and in magnetic or dielectric modulators for application in a variety of countermeasures equipment. Examples are intercept receivers, direction finders, automatic search and lock-on equipment, and FM modulators for jamming equipment.

28.2.2 Basic Circuits for Electronic Tuning

Magnetic tuning is accomplished by varying the d-c magnetic control field in a magnetic tuning unit (Figure 28-14a). This is done by varying the cur-
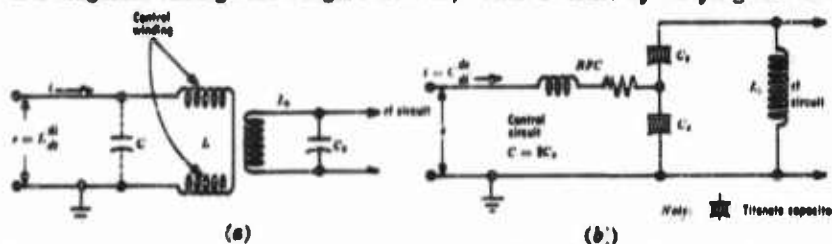
FIGURE 28-14 Magnetic and electric tuning circuits (Fig. 8, Tech. Rept. 32, EDG, The University of Michigan, Ann Arbor). (a) magnetic; (b) electric.

rent in a control winding, and electronic means must be provided to furnish the required variable current i. To obtain sufficient maximum magnetic field at a reasonable control current, a large number of turns is required on the control winding. This generally results in a control inductance of several henries. If L is the inductance of the control winding, the electronic means must furnish a voltage $L\,di/dt$ to produce the desired rate of change of current.

Electric tuning is accomplished by varying the d-c electric field in the titanate tuning capacitors (Figure 28-14b). Electronic means must be furnished to do this by changing the voltage e. To obtain sufficient maximum electric field at a reasonable voltage the dielectric in the tuning capacitors is made thin. If C is the effective capacity in the control circuit (generally 4 times the rf resonating capacity), the electronic means must furnish a charging current of $C\,de/dt$ to produce the desired rate of change of electric field.

With relatively low scanning rates (i.e., 60 cycles) there is no serious problem in furnishing the electronic means for either magnetic or electric tuning. When the scanning rate is increased, the control circuit design problems become more difficult. For high scanning rates, a short "flyback" or return time (say on the order of 1 μsec) is desired. This presents a rather

severe problem in magnetic tuning as illustrated by the following example:

A rapid-scan magnetic tuning unit tunes from 50 to 100-mc signal frequency and has the following properties:

Control inductance = 2.0 henries
Control current   = 50 ma (max.)

For a flyback time of 1.0 μsec, the flyback voltage is found to be

$$L\frac{di}{dt} = 2.0 \times 0.05 \times 10^6 = 100,000 \text{ volts}$$

This serves to illustrate the seriousness of the problem not only in control circuit design but also in voltage insulation in the design of the control winding itself.

In order to improve the situation, we first note that the magnetic field is proportional to the number of ampere turns NI. The inductance of the control winding is proportional to N^2, so the flyback voltage may be reduced at the expense of a larger control current i.

Let us re-examine the calculation just made if N is reduced by a factor of 10. We then have a unit which has the following control constants for the same maximum NI:

Control inductance = 0.02 henry
Control current   = 500 ma (max.)

For a flyback time of 1.0 μsec, the flyback voltage is found to be

$$L\frac{di}{dt} = 1.02 \times 0.5 \times 10^6 = 10,000 \text{ volts}$$

Thus an improvement is achieved in flyback voltage, but at a price of increased control current maximum, which imposes a difficult control circuit design problem.

There is another serious effect in flyback with magnetic tuning. The control inductance resonates with its self-capacity C in Figure 28-14a and generates a serious flyback transient if not properly damped. Since it is highly desirable to use pentodes for the control circuit because of their constant current property, an additional damping circuit is generally required for rapid flyback to suppress the flyback transient.

A corresponding example of rapid-scan electric tuning is a unit tuning from 50 to 100-mc signal frequency with the following properties:

Capacitance in control circuit = 200 $\mu\mu f$
Control voltage swing   = 500 volts (max.)

For a 1.0 μsec flyback time, the flyback current is found to be

$$C \frac{de}{dt} = 200 \times 10^{-12} \times 500 \times 10^6 = 0.1 \text{ ampere}$$

This flyback current may be easily obtained from a conventional vacuum tube, so the design problem is reasonably simple. Even faster flyback times are possible if thyratrons are used.

There is also no problem of flyback transients with electric tuning. It is true that the stray circuit inductance might tend to resonate with the control circuit capacity; however, this is easily damped. A series resistor is generally present for isolation purposes, so that no additional damping is required.

There are thus a number of definite advantages in electric tuning for rapid-scan applications. It may also be noted that, as the signal frequency is increased, the control capacity C generally decreases. In this case it is seen that faster flyback times are possible at higher signal frequencies without increase in the flyback current pulse. This is not the case in magnetic tuning where the control inductance does not decrease as the signal frequency is increased.

28.2.3 Magnetic Modulator Circuit
A simple balanced magnetic modulator is shown in Figure 28-15. It con-
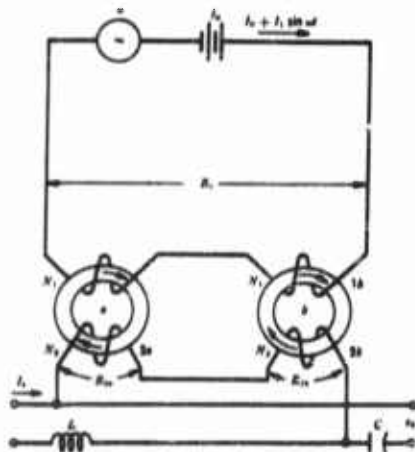
FIGURE 28-15 Magnetic modulator (*Fig. 9, Tech. Rept. 37, EDG, the University of Michigan, Ann Arbor*).
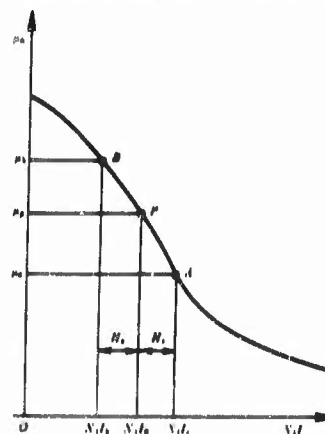
FIGURE 28-16 Characteristics used in magnetic modulator design (*from Fig. 10, Tech. Rept. 37, EDG, The University of Michigan, Ann Arbor*. The following substitutions have been made: $N_1 I_b$, $N_1 I_0$, $N_1 I_a$, $N_1 I$ for H_b, H_0, H_a, H.

sist of two similar ferrite toroid cores (*a* and *b*). These are wound with excitation windings 1*a* and 1*b* connected in series. These windings are excited by rf and d-c current generators, giving a combined current of $I_1 \sin \omega t + I_0$ amperes. A d-c current I_0, is fed through an isolating inductor L to secondary windings 2*a* and 2*b* connected in series opposition. When properly balanced, the output voltage e_o is an rf carrier having its magnitude and phase related, respectively, to the size and polarity of the signal current I_s.

The operating characteristic of each core is shown in Figure 28-16. When $I_s = 0$ both cores operate at point P by virtue of the d-c current I_0; and there e_o is equal to zero. When a positive signal current flows, core *a* operates at point *a* while core *b* operates at point *b*. Under these conditions, assuming a sinusoidal variation of flux density in the two cores (*i.e.*, B to A, a linear portion of the curve) this peak output voltage is given by

$$E_o = \omega(\mu_b - \mu_a)\, N_1\, N_2\, I_1\, G$$

where G is a factor depending on the geometry of the cores. For a linear region of the curve,

$$\mu_b - \mu_a = 2\frac{N_2}{N_1}\, I_s\left(\frac{d\mu_a}{dI_0}\right)$$

$$E_o = 2\omega\mu_{\Delta\rho}\, N_1^2 I_1\, G$$

so that the transfer impedance is given by

$$Z_1 = \frac{E_o}{I_s} = \frac{E_1 N_2^2}{\mu_{\Delta\rho} N_1^2}\left(\frac{d\mu_a}{dI_0}\right)_s$$

Note that the transfer impedance is independent of the geometry but a function of E_1 and the μ_Δ characteristic.

28.2.4 Time-Relaxation Phenomena

Time-relaxation effects are exhibited by both ferrites and titanate ceramics. In the case of ferrites a time decrease of permeability is observed after a ferrite core is subject to a degaussing treatment. When used as a tuning element in a low-power oscillator, this effect can cause up to 3.8 percent increase in frequency in ferramic G over a period of several hours. This effect was reported as early as 1947, and is variable from one material to another. In general, those materials showing a large change in μ_Δ with applied H, have also a relatively large relaxation effect.

A similar relaxation effect is noted in titanate ceramics, although this does not correspond exactly to the effect just described. If an electric field is applied to a specimen of titanate ceramic for a period of several minutes and then suddenly removed, the dielectric constant increases with time, first abruptly and then slowly. When such a ceramic is used as a tuning element

In a low-power oscillator, the frequency drops abruptly at the time the field is removed. This is followed by a relatively slow decrease in frequency. The frequency drift substantially ceases after a period of 5 to 30 minutes, depending on the material. In this respect the effect is shorter lived than the corresponding magnetic relaxation. In general, the longer period is associated with materials which show a large change in ϵ_A with applied electric field.

The consequences of these effects in swept receiver applications are given below.

28.2.4.1 Reduction of Tuning Range. When swept at 60 cycles or faster, the tuning range will be smaller by several percent than it is with manual, arbitrarily slow tuning.

28.2.4.2 Frequency drift. If the amplitude of frequency sweep is reduced, or stopped altogether, there will be a slow frequency drift. The drift will probably be small in both magnetic and electric tuning, but is shorter lived in the case of electric tuning.

28.2.4.3 Sweep Waveform. If a sawtooth sweep is employed, it is more satisfactory for electric tuning to use a jump-rise, slow-decay wave than a jump-drop, slow-rise waveform. The reason for this is that the ferroelectric ceramic responds more quickly to a jump-rise step-function of voltage than to a jump-drop step-function. A consequence of this is that the ferroelectric-tuned unit is best used to sweep downward in frequency, starting at the highest frequency and sweeping to the lowest.

In magnetic tuning, the frequency may be swept in either direction, but there is some slight preference for sweeping upward in frequency because of circuit considerations. This mode of sweeping requires a slow rise and abrupt drop in control current. Because the control tube may be rapidly cut off, permitting a high inductive kick from the control winding, the flyback time may be made somewhat shorter than in the other mode.

28.8 Geometry of Tuning Elements

Several configurations have been utilized for magnetic tuning elements. The simplest is a toroid with a bias or control winding and a signal winding. This is in general unsatisfactory because of coupling between the two windings.

The slotted core configuration is shown in Figure 28-17. A typical tuning characteristic for an oscillator circuit using this construction is shown in Figure 28-18.

Another structure that has given satisfactory performance is the resonant coaxial line shown in Figure 28-19. Structure of the tuning unit is shown in Figure 28-19a and installation in the yoke in Figure 28-19b. Typical data for an oscillator using this form of tuning unit is given in Table 28-I (page 28-59). Data for various values of tank circuit capacitance are given.

A curve showing hysteresis for this unit is given in Figure 28-20.
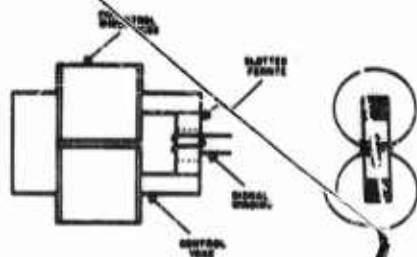
FIGURE 28-17 Typical slotted configuration driven by a yoke.
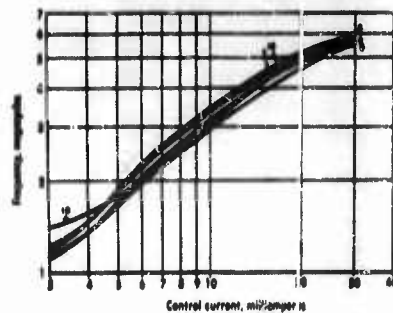
FIGURE 28-18 Typical tuning characteristics for an oscillator circuit. Typical slotted unit shows curves for several selected cores of general ceramic and body. Illotted with general ceramic G body frequency vs. control. Current from demagnetized point units made from preselected cores

FIGURE 28-19 Resonant line-tuning unit.
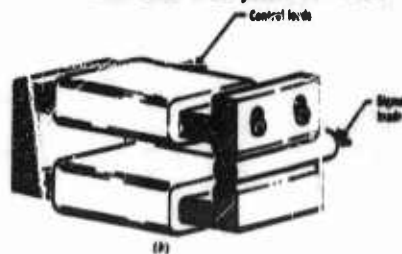
FIGURE 28-20 Tuning curve for resonant-line unit showing hysteresis.

FIGURE 28-21 Capacitor tuning units.
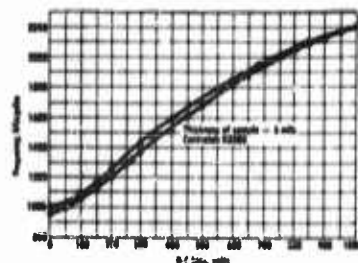
FIGURE 28-22 Ferroelectric oscillator tuning curve (*Fig. 35, Tech. Rept. 31, EDG, The University of Michigan, Ann Arbor*).
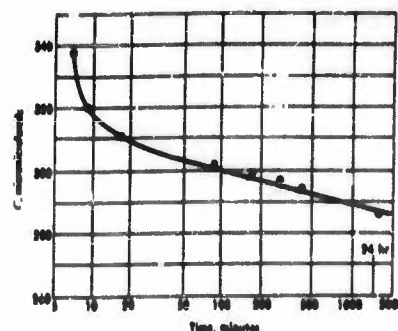
FIGURE 28-23 Temperature effect on ferroelectric tuning (*Fig. 8, Tech. Rept. 31, EDG, The University of Michigan, Ann Arbor*). *C* vs time. Ageing after heating to 100°C and quenching at 25°C, Aerovox "HI-Q" body No. 41.
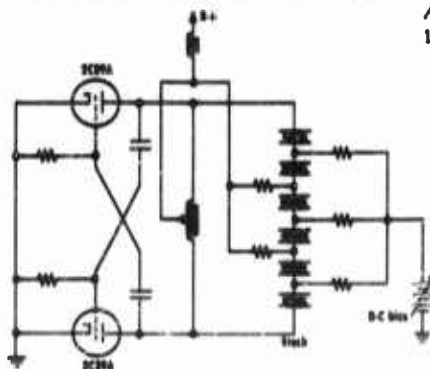
FIGURE 28-24 High-power oscillator circuit. (*Proceedings of the National Electronics Conference*, 1955, *p.* 846). Voltage-tunable power oscillators.
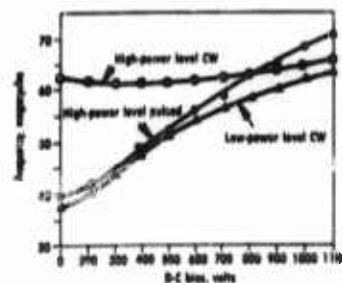
FIGURE 28-25 Power oscillator tuning characteristics (*Proceedings of the National Electronics Conference*, 1955, *p.* 847).

The structure of low- and high-power (stack) ferroelectric capacitor tuning units is shown in Figure 28-21. These capacitors must be suitably encapsulated to keep out moisture.

A typical oscillator tuning curve is shown in Figure 28-22. The hysteresis effect is evident. The effect of temperature on tuning is shown in Figure 28-23.

In Figure 28-24 a high-power oscillator circuit is shown. In this case the capacitor was built in a stack with radiating fins between elements. The effect of self-heating of the capacitor is indicated in Figure 28-25. At high power levels the tuning is almost eliminated because of change in ceramic

characteristics with temperature. This oscillator was operating at about 3-watt level for the high power level and at about 100 milliwatts for the low power level.

28.4 Electronically Tuned Intercept Receivers

With the basic tuning elements just described it is possible to construct scanning receivers with a variety of characteristics. Details of the design problems will not be covered here. It is evident from what has been discussed that further research is indicated to minimize problems in the following areas:

(1.) Temperature (through selection of materials, thermostatic oven control, or use of afc).

(2.) Hysteresis effects and flyback transients (through design of suitable sweep circuits).

(3.) Tracking problems in superheterodyne receivers (through material selection and circuit design).

(4.) Control power, voltages, and currents.

(5.) Time-decay (through crystal control or other stabilizing circuits).

Advantages of electronic tuning are in the possibility of increased scanning rates, reduction of weight, size, and cost, and elimination of other disadvantages of mechanically tuned receivers.

Receivers have been constructed for use in the frequency range from 50 kc to somewhat less than 500 mc. The number of tuning heads varies with the specific design. It is possible to cover the range 50 mc to 500 mc with 4 or 5 heads. Below this frequency about 2 or 3 heads for each decade of frequency is required.

Expected improvements in materials should make the application of these materials more widely accepted and used.

II. Microwave Applications of Ferrimagnetic Materials

One of the significant recent advances in microwave technology has been the achievement of nonreciprocal and electrically controllable circuit elements through use of the properties of special materials. One area of application utilizes the properties of ferrites and ferrimagnetic garnets under the influence of static magnetic fields. These media derive their useful microwave properties primarily from interactions between electrons in the media and incident microwave signals. These interactions are strongest when the frequency of the microwave signal coincides with the natural resonant frequency of the electrons, which in turn is directly proportional to the effective applied magnetic field. The fact that a gross material property such as permeability can be controlled by means of an applied magnetic field, and that it can be used to give both reciprocal and nonreciprocal microwave effects, makes these

materials of special interest for microwave applications.

The following sections review some of the fundamental characteristics of ferrimagnetic materials and describe the utilization of these characteristics for device applications, with particular emphasis to some of the special problems of countermeasures.

28.5 Fundamental Characteristics of Magnetic Ferrites

28.5.1 Structure

Ferrites are nonmetallic magnetic materials. They are made of a hard ceramic with a crystalline structure similar to that of the mineral spinel, $MgAl_2O_4$. Most ferrites can be expressed by the simple chemical formula XFe_2O_4, where X is a bivalent metallic ion. The oxygen ions have radii of 1.4 Å, which is roughly twice that of the X^{++} and Fe^{+++} ions. Most acceptable metal ions have radii of 0.6 to 1.0 Å. Commonly used metal ions are those of magnesium, manganese, and nickel.

In a ferrite, the large oxygen ions form a face-centered cubic structure of spheres. The small metal ions are located in the resultant interstices. If the oxygen geometry is examined, it is found that two kinds of interstices exist. One interstice is surrounded by 4 oxygen ions and is called a tetrahedral site. The other interstice is surrounded by 6 oxygen atoms and is called an octahedral site. A unit ferrite crystal cell consists of a lattice of 32 oxygen ions containing 96 interstices. Of those, 64 are tetrahedral sites and 32 are octahedral sites. The 24 metal ions in a unit cell of the spinel structure are distributed with the 16 trivalent ions in octahedral sites, while the 8 divalent ions are found in tetrahedral sites. There is a second common ferrite structure in which the 8 divalent metal ions replace 8 of the iron ions in the octahedral sites. The 8 displaced trivalent iron ions then occupy tetrahedral sites. This distribution is called the inverse spinel structure.

The ferrimagnetic garnets are formed from oxides of rare earth metals like yttrium and gadolinium. The chemical composition is $3Y_2O_3 \cdot 5Fe_2O_3$. These materials have properties very much like the ferrites and, because of narrower intrinsic line width, are superior for some applications.

28.5.2 Magnetic Properties

Fundamentally, the magnetic properties of every material is determined by the behavior of the electrons it contains, since the nuclear contribution is negligibly small in comparison. The magnetic effects are due primarily to the spins of the electrons and to a much smaller degree to their motions in atomic orbits. In all atoms the electron spins tend to pair off against each other so that the net magnetic moment of the atom is reduced to that of a few of its

electrons, and in many atoms the spins are all paired against opposing spins so that the atom has no net magnetic moment due to spinning electrons. The latter are the atoms which are diamagnetic, having only the magnetic effects due to orbital motions. Ferromagnetic materials are composed of atoms having permanent magnetic moments due primarily to unneutralised electron spins. In addition, very large internal forces cause neighboring dipoles to line up parallel to one another, a condition of self-saturation.

The domain theory of magr   ' on of a ferromagnetic material assumes that the material is composed o. .. ny small regions or domains, each magnetized to saturation in some direction. In the absence of an applied magnetic field, the directions of saturation of the domains are in general distributed in a way such that the resultant magnetization of the specimen as a whole is zero.

In the usual ferromagnetic domain the magnetic moments are all in parallel alignment. In the ferrite, however, ions in the octahedral site have magnetic moments that are usually aligned antiparallel to those in the tetrahedral sites. In general, the intensity of magnetization will be different for the two sites and the total magnetization is a combination of these two sites. This principle, which is called ferrimagnetism, is illustrated in Figure 28-26. This
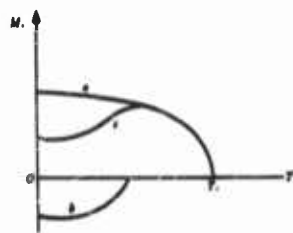
FIGURE 28-26 Temperature dependence of saturation magnetization for some ferrimagnetic materials. (a) Contribution of A-site ions; (b) contribution of B-site ions; (c) resultant magnetization.
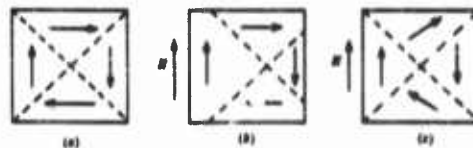
FIGURE 28-27 Magnetization processes. (a) Unmagnetized; (b) magnetized by domain growth or boundary displacement; (c) magnetised by domain rotation.

magnetization curve is but one of eight that has been predicted from this type of model. The temperature at which the magnetism goes to zero is called the Curie temperature.

The magnetic moment of any one domain is specified by the magnitude and direction of its magnetization and by its volume. When a field is applied, the moment of a domain, and therefore the magnetization of the ferrimagnetic

material of which it is a part, is ordinarily changed by (a) the growth of domains oriented favorably with respect to the field direction at the expense of domains less favorably oriented (domain wall displacement), or (b) a change in the magnetization direction of the domain by rotation toward the applied field direction (domain rotation). These conditions are illustrated in Figure 28-27.

28.5.3 Losses in Ferrites

Ferrites absorb energy from an electromagnetic radiation field by interacting with both the electric and the magnetic fields. The losses due to interaction with the electric field are generally known as dielectric losses, while those resulting from interaction with the magnetic field are called magnetic losses. Both the dielectric and the magnetic losses can be accentuated by dimensional resonances which are common at microwave frequencies because of the large effective dielectric constant of the material.

In many ferrites the dielectric loss tangent, or the ratio of the imaginary to the real component of a complex dielectric constant, is small. However, this is not always the case, and a low d-c conductivity is not a sufficient criterion to impose for low microwave dielectric losses. Some ferrites appear to contain highly conducting regions isolated by an insulating matrix and exhibit high microwave dielectric losses. Conductivity measured at audio frequencies will show a strong frequency dependence if the material is of this type. The conducting islands act as plates of a condenser separated by the grain dimensions. Although the d-c conductivity within a grain may be high, the measured values across grain boundaries will often be low.

Magnetic losses in ferrites with no applied static magnetizing field generally occur in two regions. One occurs in the rf range under 500 mc and is due to domain wall displacements; the other occurs in the microwave range and is due to domain rotations. Both of these phenomena generally occur over a wide frequency range and thus constitute large regions of high magnetic loss. The losses due to domain rotations constitute the major microwave loss present when there is no applied static magnetic field, or when the applied magnetic field is small. They can be explained in terms of a ferromagnetic resonance with the effective internal field in the ferrite, which exists even in the absence of an applied field. The upper frequency at which this resonance occurs can be reduced by lowering the saturation moment of the material.

Almost all microwave applications of ferrites are based on either the loss or dispersive properties due to ferrimagnetic resonance under the influence of an applied static magnetic field. Accordingly, this phenomenon is worthy of special discussion.

28.5.4 Ferrimagnetic Resonance

When a ferrite is magnetized, the magnetic moments of the electrons precess about the effective stationary internal magnetic field. If a rotating rf magnetic field is applied in the opposite direction of rotation to that of the precession, there is little interaction between the field and the electrons. However, a field which is rotating in the same direction as the electrons has a stronger coupling because the field vector and the spin vector are parallel for a longer period during each revolution. As the frequency of the field rotation approaches that of the electron precession, the two vectors remain parallel, and the electron is able to absorb energy continuously from the field. This is the condition for ferrimagnetic resonance.

A simple mechanical analog to the precessing electrons responsible for ferrimagnetic resonance is shown in Figure 28-28. The spinning top precesses
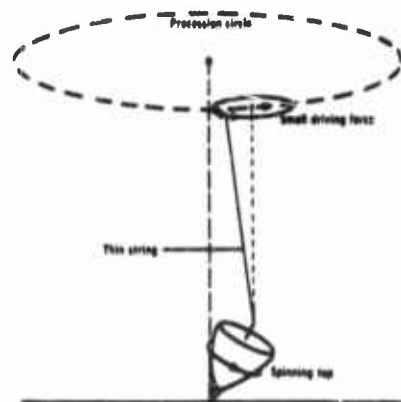
FIGURE 28-28   Mechanical analog to an electron precessing in a magnetic field.

about the earth's gravitational field. If a thin string is attached to a point on the axis of the top, energy can readily be transmitted to the top by rotating the other end of the string in a small circle whose center follows the projection of the precession path of the top. If the end of the string is circled in the same direction as the top is precessing, a maximum amount of energy is transmitted to the top when these two angular frequencies are the same. Under these conditions the system is in resonance. If the end of the string is rotated in the opposite direction, very little energy is transmitted to the top.

In the model described, the top is analogous to an electron of a ferrimagnetic medium and the earth's field to the magnetic field required for precession. The driving force at the end of the string is analogous to a force on an

electron arising from interaction with the rotating magnetic field of a circularly polarized rf wave. The frequency at which the electrons precess about a stationary magnetic field is directly proportional to the magnitude of that field. To observe resonance one has the choice of varying either the operating frequency or an applied field until the operating frequency equals the precession frequency of the electrons.

28.5.5 Permeability Tensor

When a microwave magnetic field is present in a ferrite that has a static magnetic field applied to it and the microwave field is in any direction other than along the static field, then the electrons precessing about the static magnetic-field direction will interact with the microwave field. They couple energy out of the microwave field and deliver it back, but part of it is 90° out of time and space phase with the driving microwave field. Thus, the relationship between the magnetization and the inducing microwave magnetic field is not the usual scalar type of relation but rather a tensor type. This means that the magnetization in a given direction depends not only on the inducing field in that direction but upon the field in other directions as well.

The motion of the magnetization for the lossless case is given by the torque equation

$$\frac{dM}{dt} = \gamma(M \times H) \qquad (28\text{-}1)$$

where $H = H_0 + h$ = internal magnetic field

$M = M_s + m$ = magnetization of medium

H_0 = static magnetic field in s direction

h = a-c magnetic field

M_s = saturation magnetization in s direction

m = a-c magnetization of medium

γ = gyromagnetic ratio of the electron

$= -2.8$ mc/oersted or -2π (2.8) megaradians/oersted

The mathematics is simplified if a coordinate axis is chosen along the direction of the static magnetic field H_0. H_0 is assumed to be large enough to magnetically saturate the ferrite and to be much larger than the rf magnetic-field components. Harmonic time dependence is assumed for the alternating magnetic field and magnetization, and terms of second order in small quantities are neglected. Then the components of Eq. (1) are:

$$j\omega m_s = \gamma H_0 m_y - \gamma M_s h_y$$
$$j\omega m_y = \gamma M_s h_s - \gamma H_0 m_s \qquad (28\text{-}2)$$
$$j\omega m_s = 0$$

Solving for the magnetizations in the x and y directions gives

$$4\pi m_x = \chi h_x - j\kappa h_y$$
$$4\pi m_y = j\kappa h_x + \chi h_y \tag{28-3}$$

where

$$\chi = \frac{4\pi M_s \gamma^2 H_0}{\gamma^2 H_0^2 - \omega^2} \tag{28-4}$$

and

$$\kappa = \frac{4\pi M_s \gamma \omega}{\gamma^2 H_0^2 - \omega^2} \tag{28-5}$$

It can be seen from Eq. (4) and (5) that resonances will occur whenever $\omega = |\gamma| H_0$, which is the gyromagnetic resonance condition. Loss has been neglected here; including it makes both χ and κ complex instead of real.

The permeability for the infinite medium is a tensor* relating the flux density and magnetic field

$$\mathbf{b} = [\mu]\mathbf{h} \tag{28-6}$$

where

$$[\mu] = \begin{pmatrix} \mu & -j\kappa & 0 \\ j\kappa & \mu & 0 \\ 0 & 0 & \mu_0 \end{pmatrix} \tag{28-7}$$

and

$$[\mu] = 1 + \chi$$

It is this infinite-medium permeability tensor which must be used in Maxwell's equations for regions containing ferrites. It is assumed in Eq. (7) that the static magnetic field is along the z direction.

28.5.6 Microwave Propagation in Infinite Medium

The propagation conditions for a plane wave propagating in an infinite ferrite medium are determined from the solution to the vector wave equation

$$\nabla \times \nabla \times h - \beta_0^2 \epsilon_r [\mu] h = 0 \tag{28-8}$$

where β_0 = free-space propagation constant
   ϵ_r = relative dielectric constant
   $[\mu]$ = permeability tensor given by Eq. (7)

Two special cases are considered below.

*Tensors are indicated by brackets.

28.5.6.1 *Magnetic Field Parallel to Propagation.* Microwave applications of ferrites were first developed with propagation parallel to the applied magnetic field. For this case the two solutions to the wave equation for the anisotropic medium represent counterrotating circularly polarized waves. One of the waves is rotating in the direction for maximum interaction with the precessing electrons of the magnetized medium, while the other wave is rotating in the direction for minimum interaction. Each direction of polarization is characterized by a different effective scalar propagation constant.

$$\beta_{\pm}^{2} = \beta_0^{2}\epsilon_r(\mu \mp \kappa) \qquad (28\text{-}9)$$

The plus and minus subscripts correspond to right- and left-hand circularly polarized waves.

The effective scalar permeability for the two waves is

$$\mu_{\pm} = \mu \mp \kappa = 1 + \frac{4\pi M_s \gamma}{\gamma H_0 \pm \omega} \qquad (28\text{-}10)$$
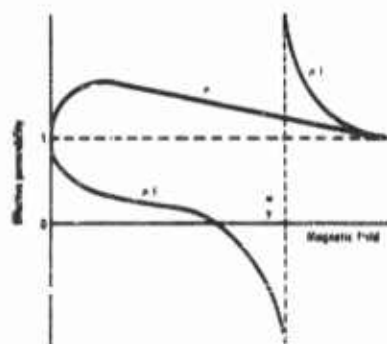
FIGURE 28-29 Effective permeabilities for right- and left-hand circularly polarized waves in infinite medium magnetized parallel to direction of wave propagation.

where γ is a negative number. μ_+ is the effective permeability for the wave that interacts strongly with the precessing electrons and, as shown in Figure 28-29, resonance occurs when $H_0 = \omega/|\gamma|$. Thus it is seen that infinite-medium plane-wave theory indicates that the two circular polarizations are the normal modes leading to scalar permeabilities in the wave equations.

28.5.6.2 *Magnetic Field Perpendicular To Propagation.* An infinite ferrite medium magnetized perpendicular to the direction of wave propagation behaves like an isotropic dielectric if the rf magnetic field is parallel to the direction of magnetization. If the rf magnetic field is normal to the direction of magnetization, different propagation characteristics are exhibited. Solution to the wave equation gives the following for the propagation constants:

$$\beta_{\parallel}^{2} = \beta_0^{2}\epsilon_r \qquad (28\text{-}11)$$

$$\beta_\perp{}^2 = \beta_0{}^2 \epsilon_r (\mu^2 - \kappa^2)/\mu \qquad (28\text{-}12)$$

An effective permeability can be ascribed to each of the two waves. For an rf field parallel to the direction of magnetization the effective permeability is

$$\mu_{||} = 1 \qquad (28\text{-}13)$$

If the two are perpendicular, the effective permeability becomes

$$\mu_\perp = \frac{(\mu^2 - \kappa^2)}{\mu} = \left[\frac{\gamma^2 B^2 - \omega^2}{\gamma^2 H_0 B - \omega^2} \right] \qquad (28\text{-}14)$$

A plot of the two effective permeabilities is shown in Figure 28-30. Resonance occurs for the perpendicular wave when $\omega = (H_0 B)^{\frac{1}{2}}/|\gamma|$.

28.5.7 Kittel's Equation for Bounded Medium

The previous relationships apply for propagation in an infinite ferrite medium. In using the results for practical applications, it is important to understand the effects of the ferrite shape in determining the microwave characteristics. When a ferrite body is placed in an applied magnetic field, a magnetization is induced within the body. This magnetization terminates on the surface of the ferrite body, creating magnetic poles on the surface. These poles are the source of a field within the ferrite, commonly called the demagnetizing field. It is opposed to the applied field. The proportionality constant between the demagnetizing field and the magnetization M is called the demagnetizing factor N. The demagnetizing factor is thus a measure of the effectiveness of the poles at the surface of a ferrite body in producing a field inside of the body.
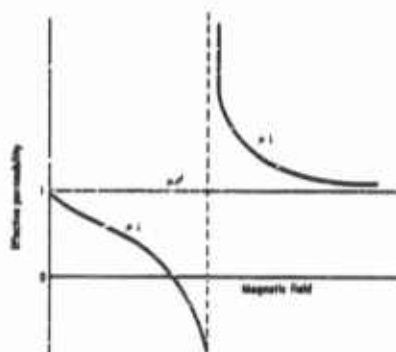
FIGURE 28-30 Effective permeabilities for the infinite medium magnetized normal to direction of wave propagation.

It can be shown that the sum of the demagnetizing factors for any three orthogonal directions is 4π.* For example, in a long rod magnetized along its axis, the poles are widely separated, the induced demagnetizing field is very

*For magnetic induction defined by $B = \mu_0 (H + 4\pi M)$.

weak, and thus the demagnetizing factor along its axis is zero. A thin disk magnetized along its axis has very close poles and a large demagnetizing field. Hence the demagnetizing factor along the axis of the disk is 4π. These examples are the extremes; for other shapes the demagnetizing factor for a given direction is between 0 and 4π.

The relation between the internal field H_{int}, the applied field H_{app}, and the demagnetizing field NM in a given direction is

$$H_{int} = H_{app} - NM \qquad (28\text{-}15)$$

This equation is exactly valid only for the case of an ellipsoid since only then is the demagnetizing factor a constant. However, this relation gives reasonably good results for shapes such as disks and rods which are only approximately ellipsoidal. If Eq. (15) is written for the three directions and these internal fields substituted into the magnetic torque equation, a relation between the ferromagnetic resonant frequency ω_{res} and the applied fields is found. This relation is commonly called Kittel's equation. It can be written formally as

$$\omega_{res} = |\gamma| H_{eff} \qquad (28\text{-}16)$$

where the effective magnetic field H_{eff} is given by

$$H_{eff} = \{[H_{app} - (N_z - N_x)M_s][H_{app} - (N_z - N_y)M_s]\}^{\frac{1}{2}} \quad (28\text{-}17)$$
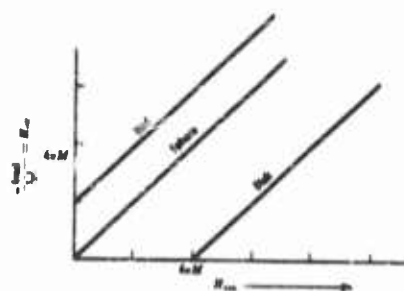
FIGURE 28-31 Ferromagnetic resonance frequency as a function of applied magnetic field, illustrating the shape effects for a sphere and two extremes of a rod with length much greater than diameter and a disk with diameter much greater than thickness. H_{app} is along axis of both rod and disk.

and the applied field is in the z direction. Figure 28-31 illustrates several cases of Kittel's equation and shows that the effective magnetic field can be either greater or less than the applied field, depending on whether the ferrite sample is more rodlike or more disklike.

An additional fact which can lead to confusion is that the terms above, at, and below the resonant frequency are synonymous, respectively, with below, at, and above the resonant field. This can be shown by the following argument. Suppose a ferrite is excited at frequency ω (actual) and has a magnetic field applied to it which deter-

mines an effective field H_{eff} (actual). Assume that the field H_{eff} (actual) is not the field for resonance at the frequency ω (actual). The effective field necessary for a resonance at ω (actual) is H_{eff} (res). The resonant frequency associated with H_{eff} (actual) is ω (res). In equation form,

$$\omega \text{ (actual)} = |\gamma| H_{eff} \text{ (res)} \text{ and } \omega \text{ (res)} = |\gamma| H_{eff} \text{ (actual)} \qquad (28\text{-}18)$$

Thus if

$$\omega \text{ (actual)} > \omega \text{ (res)} \qquad (28\text{-}19)$$

then

$$H_{eff} \text{ (actual)} < H_{eff} \text{ (res)} \qquad (28\text{-}20)$$

The two inequalities show that operation at a frequency above the resonant frequency is synonymous with operation at a field below the resonant field.

To review: the following is a listing of the various fields that have been defined.

H_{app} = the applied static magnetic field from external sources

H_{int} = the magnetic field inside the body

M = magnetization or magnetic moment per unit volume

  (M_s = magnetization at saturation, when all the domains are aligned in a single direction)

$N_i M_i$ = demagnetizing field in the ith direction

H_{eff} = effective magnetic field, a fictitious field defined by Eq. (17)

28.6 Ferrite Microwave Device Principles

Ferrite microwave devices make use of the dispersion characteristics of the ferrite permeability due to interactions of electrons in the ferrite with the incident microwave signals. As shown in Section 28.5.6, the normal modes of propagation are circularly polarized waves. Strong interaction occurs for an electromagnetic wave circularly polarized in the same direction as that in which the electrons are precessing (plus direction), while only weak interaction occurs for a wave circularly polarized in the opposite (minus) direction. Accordingly, for device applications, a piece of ferrite generally is placed in a region of circular polarization, or for certain applications in a region of linear polarization, which can be decomposed into its two circular components.

The shape of the dispersion curve as a function of magnetic field leads to three general principles of device operation: resonance absorption, field displacement, and differential phase shift or Faraday rotation. This is illus-
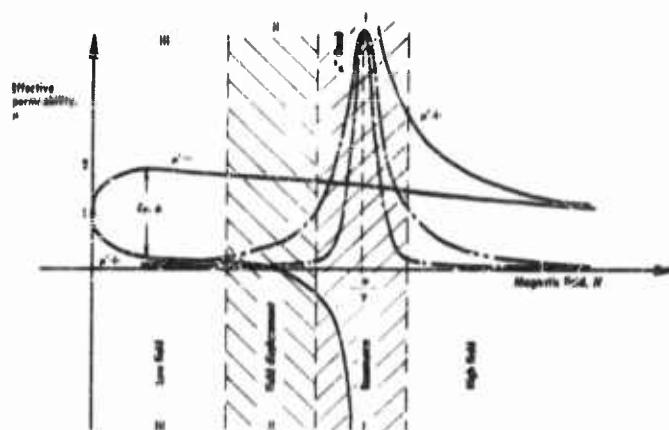
Figure 28-32 Effective permeabilities for right- (+) and left- (−) hand circularly polarized waves. Propagation parallel to magnetic field, showing magnetic field bias required for various device principles.

trated schematically in Figure 28-32 for the case of propagation parallel to the applied magnetic field. For applications based on resonance absorption (region I), the effective applied magnetic field is adjusted until the electron precession frequency equals the frequency of the microwave signal. Then for circularly polarized waves there is a large energy absorption while for negative circularly polarized waves the loss remains low. Application in region II are based on the fact that plus waves are largely excluded from the ferrite, because the propagation constant of the ferrite is approximately zero, while minus waves can propagate through the ferrite in normal fashion. Applications in region III are possible because differential phase shift or Faraday rotation depends on the difference between the two permeability curves.

Devices based on these three principles are discussed in more detail below. Applications based on longitudinal and transverse magnetic field are discussed separately.

28.6.1 Applied Magnetic Field Parallel to Direction of Propagation

28.6.1.1 Faraday Rotation Applications. In 1845 Faraday passed light through a foil magnetized in the direction of propagation and observed a rotation of the plane of polarization. He also observed that the angle of rotation, proportional to the thickness of material traversed, is a function of the applied magnetic field and is in the same direction whether the wave is

propagated in the direction of the magnetic field or opposite to it. These same phenomena have been observed in the microwave region and are in qualitative agreement with the predicted results.

The relations given in Section 28.5.6 are of special interest because any linearly polarized wave may be resolved into two counterrotating circularly polarized waves providing compatible boundary conditions exist. Equation (9) shows that the two circular components of a linear wave are propagated with different propagation constants corresponding to different phase velocities. Upon emerging from the anisotropic medium and recombining, the two circular waves result in a linear wave rotated relative to its initial orientation. Since $\beta_+ \neq \beta_-$, it is apparent that the resultant of the two circular waves is rotating in the direction of the one with the larger phase factor. The effective phase factor is $(\beta_+ + \beta_-)/2$, and the plane of polarization is rotated by an angle

$$\theta_F = (\beta_- - \beta_+)\, s/2 \qquad (28\text{-}21)$$

This rotation is known as Faraday rotation. Substituting from Eq. (9) and (10) into Eq. (21) yields

$$\theta_F = \frac{\omega \epsilon^{\frac{1}{2}}}{2c}\left[\left(1 + \frac{4\pi M_s \gamma}{\gamma H_0 - \omega}\right)^{\frac{1}{2}} - \left(1 + \frac{4\pi M_s \gamma}{\gamma H_0 + \omega}\right)^{\frac{1}{2}}\right] \qquad (28\text{-}22)$$

The magnetic-field dependence of the Faraday rotation is illustrated in Figure 28-33. The rotation is roughly linear with field up to the point of magnetic saturation, which normally occurs for applied fields under 100 oersteds. The rotation then remains essentially constant until field values approach the region of ferromagnetic resonance.

There exist a large number of microwave devices based on Faraday rotation. Some of these depend on a fixed rotation, and others depend on variations in rotation controlled by changes in the applied magnetic field.

FIGURE 28-33 Dependence of Faraday rotation on magnetic field at constant frequency.

Faraday rotation devices are generally constructed with ferrite rods mounted coaxially in circular or square waveguides or any structure having fourfold rotational symmetry. The rod is magnetized along its axis in one of several different ways. For variable magnetic fields, a solenoid is generally

wound about the waveguide containing the ferrite rod. Fixed longitudinal fields are achieved by placing small permanent rod magnets inside the waveguide between two ferrite rods, or by placing ring magnets around the outside of the waveguide; the small rod magnets introduce undesirable frequency sensitivities for certain applications.

Isolator. One of the first recognized applications of magnetic ferrite microwave properties was that of a one-way transmission system, or a load isolator, using Faraday rotation. The manner in which this is achieved is illustrated in Figure 28-34. The ferrite sample length and the biasing magnetic
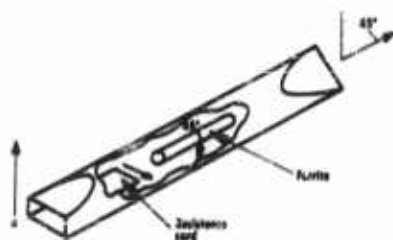
FIGURE 28-34 Faraday rotation micro-wave isolator.
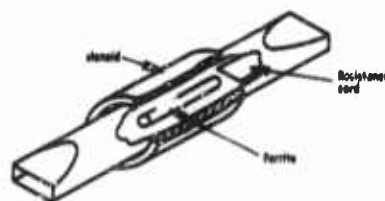
FIGURE 28-35 Microwave switch or amplitude modulator.

field are adjusted to give 45° Faraday rotation. Then, energy entering the generator end polarized along *a* is rotated 45° by the ferrite to *b*. This polarization is acceptable to the output waveguide; hence there is transmission with low loss. Energy reflected from the load enters the rotator polarized along *b* and is rotated 45° by the ferrite into the direction *c* where it is absorbed by a resistance card. Thus, the nonreciprocity is due to the fact that the ferrite rotates the plane of polarization in the same direction in space independent of the direction of propagation.

Microwave Switch or Amplitude Modulator. Another possible application is suggested by Figure 28-35. Here the magnetic field can be varied so that the Faraday rotation varies from 0° to 90°. For zero rotation, maximum energy will be transmitted, and for 90° rotation negligible energy will be transmitted. Varying the magnetic field produces an on-off switch or provides amplitude modulation. Typical characteristics for this type of switch are maximum and minimum attenuations of approximately 30 db and 0.25 db, respectively, and a d-c power requirement of about 1 watt.

Microwave Gyrator. A gyrator is an element for which the impedance matrix off-diagonal elements Z_{ij} are the negative of Z_{ji}. The gyrator is similar to the device shown in Figure 28-34, except that the waveguide twist and the Faraday plate each give 90° rotation. A wave passing from left to right receives 180° rotation or 180° phase shift, while one passing from right to left

receives none. The gyrator is a fifth net work element which can be used as a basic circuit element for network synthesis. It is comparable in theoretical significance to the capacitor, resistor, inductor, and transformer.

Microwave Circulator. The circulator is a multiple port device designed to transmit a signal from one of its ports to an adjacent port while isolating the signal from all other ports; that is, a signal into Port One goes out Port Two; a signal into Port Two goes out Port Three, etc.; and a signal into the last port goes out Port One. The usual Faraday rotation circulator has four ports, each with polarization orientation as shown in Fig. 28-36. In the center section is . . Faraday plate. Incident power with orientation a is turned into polarization b. Similarly, b is turned into c, c into d and d into a.
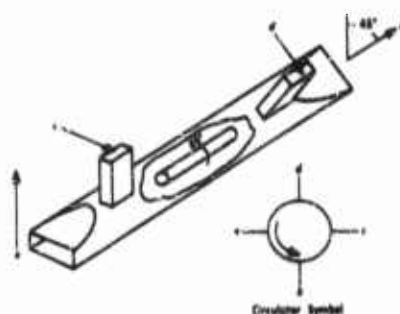
FIGURE 28-36 Faraday rotation microwave circulator and schematic symbol for circulator.

Feedback Amplitude Stabilizer. A device similar to the amplitude modulator can be used as a control element to give constant amplitude from a sweep oscillator. A sensing element controls the current to the rotator solenoid in such a manner as to counteract changes in output amplitude.

Polarization Detection and Modulation. Faraday rotation is useful for rapidly determining the orientation of the linear TE_{11} wave in a circular waveguide. A detector located in a predetermined direction shows a maximum output when the signal is rotated into that plane by a Faraday rotator. A rotator also can be used to modulate the polarization of a transmitted signal.

28.6.1.2 Microwave Energy Displacement Applications. A large difference in the energy distribution can be obtained for the two counterrotating TE_{11} modes in a circular waveguide containing a ferrite rod. Equation (10) and Figure 28-32 indicate that a ferrite rod magnetized along its axis presents an effective rf permeability of between 1 and 2 for a negative circuit polarized wave. For the positive wave, the field can be adjusted so that the effective rf permeability becomes very small and negligible energy is transmitted through the ferrite rod. With the large dielectric constant ϵ_r normally possessed by ferrites, the product $\mu \epsilon_r$ may be sufficiently large to permit the negative wave to propagate within the ferrite rod as in a dielectric waveguide.

Energy displacement due to a magnetized ferrite also can be used to make

isolators, microwave switches, or amplitude modulators. The microwave geometry used is similar to that shown in Figure 28-37.
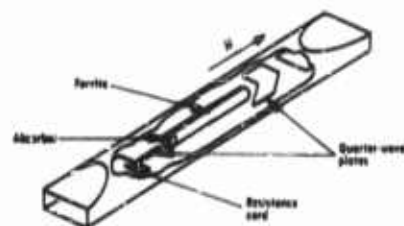
FIGURE 28-37 Energy displacement isolator.

Energy Displacement Isolator. A unidirectional transmission line can be constructed by adding quarter-wave plates before and after a ferrite rod with an absorber inserted in it. For the forward direction of propagation, for which the quarter wave plate converts a linear wave to one having positive circular polarization, the wave tends to go around the ferrite with small insertion loss. For the backward direction of propagation, the negative circularly polarized wave tends to concentrate inside the ferrite and is absorbed, giving a high isolation. The absorber also may be coated on the surface of the rod with similar results.

Microwave Switch or Amplitude Modulator. Amplitude modulation can be achieved by applying an alternating magnetic field to the basic isolator described in the preceding section. If the applied field is initially in the direction of maximum energy transmission and is then reversed, the microwave output can be reduced by 50 db. As the applied field swings back and forth, amplitude modulation of the microwave energy results.

28.5.1.3 Ferromagnetic Resonance Absorption Applications. Equation (10) shows that resonance occurs for the positive circular wave if $|\gamma| H_0 = \omega$. This is the condition for which the microwave frequency is equal to the frequency of electron precession. At this frequency the electrons extract a maximum amount of energy from the wave. Although Eq. (10) was derived for the loss-free medium, the resonance condition is essentially the same when the loss term is introduced. Under these conditions, as shown in Figure 28-38, the absorption can be increased or decreased by control of the magnetic field in the neighborhood of resonance.

Resonance absorption can be used as the basis for isolators, modulators, or switches and tunable filters. If the longitudinal magnetic field is adjusted for the resonance condition, the positive circular wave will be absorbed by the ferrite and the negative wave will not be appreciably affected if the ferrite is in the form of a small-diameter rod.

Resonance Absorption Isolator. One geometry for achieving a resonance-absorption isolator is similar to that shown in Figure 28-37. No absorber is required in the ferrite for this application. A quarter-wave plate placed in

$M_1 = \frac{\gamma}{\gamma}$

$H_2 = \frac{\gamma}{\gamma}\sqrt{1 + \left(\frac{4\pi M\gamma}{\omega}\right)}$
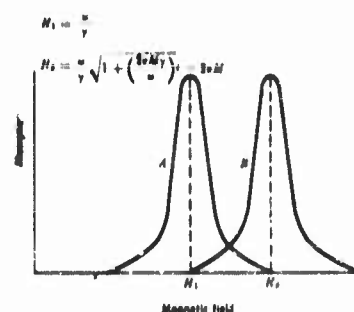
FIGURE 28-38 Ferromagnetic resonance absorption for longitudinal magnetic field (A) and transverse magnetic field (B).

front of the ferrite, so as to convert a linear input wave to negative circular polarization, will permit transmission past the ferrite element. A second quarter-wave plate placed after the element is used to convert back to the initial linear-wave orientation. If the linear wave is incident from the opposite direction, it is converted to a positive wave relative to the magnetic field and is absorbed by the ferrite.

Microwave Switch or Amplitude Modulator. If the applied magnetic field is varied about some value near the shoulder of the absorption curve shown in Figure 28-38, modulation results. Similarly, off-on switching can be obtained by rapidly changing the magnetic field from resonance to off-resonance values. Because of the high applied fields required, modulation or switching by resonance absorption for longitudinal fields is not nearly so practical as the use of Faraday rotation or energy displacement.

Coaxial and Strip-Line Tunable Filters. Ferrite-loaded coaxial and strip transmission lines provide geometries for utilizing tunable reciprocal resonance absorption. For these cases, a sleeve of ferrite is inserted into the coaxial line or a slab is inserted into the strip line and a longitudinal or transverse field is applied. The amount of attenuation produced by the element is then controllable over wide ranges by varying the amplitude of the applied field.

28.6.2 Applied Magnetic Field Transverse to Direction of Propagation

If a region of rotating rf magnetic field exists, a small piece of ferrite in this region can be used effectively to control the microwaves. Maximum interaction between the microwave energy and the precessing electrons will occur if the plane of the rotating rf magnetic field is perpendicular to the d-c field direction and if the rf field and electrons are rotating in the same sense. If they rotate with opposite senses, the interaction is very small. These field-electron interactions lead to the variations in propagation constant. The propagation constant is actually complex, containing both phase factor and attenuation, which are respectively different for the two counterrotating waves.

In a rectangular waveguide of a TE_{01} mode, there are two planes parallel

to the side walls in which the rf magnetic field is circularly polarized. Therefore, if a very thin ferrite slab is placed in one of these planes and magnetized so that the electrons precess in the direction of rf field rotation, strong field-electron interaction results. Thick ferrites, however, greatly distort the empty waveguide fields, and the field solutions must be calculated on the basis of the loaded guide configuration. Nevertheless, simplified considerations give a qualitative picture of how a transversely magnetized slab affects the waveguide transmission properties. If the d-c magnetic field were reversed, or if the slab were placed on the opposite side of the center plane, or equivalently if the direction of propagation were reversed, there would be very little interaction between the field and electrons because their directions of rotation would be opposite. These considerations form the basis for rectangular waveguide devices possessing nonreciprocal and magnetically controllable phase-shift and attenuation characteristics.

28.6.5.1 Phase-Shift Applications

Phase Modulators. Phase modulation is achieved by placing a small ferrite slab in a rectangular waveguide in a region of either linear or of rotating microwave magnetic fields, and amplitude modulating the transverse applied magnetic field. If a single slab of ferrite is placed in a region of circular polarization the phase shift will be nonreciprocal. This also is the case for two slabs biased as in Figure 28-39. However, if the two slabs are biased in the same direction the phase shift will be reciprocal.

An electronically controllable phase shifter also can be constructed by placing a ferrite in a coaxial or strip transmission line and magnetizing transversely or longitudinally with low magnetic fields. The phase shift through the element is reciprocal and is a function of the amplitude of the magnetic field.
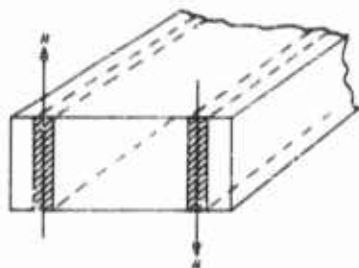
FIGURE 28-39 Nonreciprocal phase shifter. Rectangular waveguide loaded with magnetized slab of ferrite in a manner so as to make it nonreciprocal.

Circulators. Figure 28-40 illustrates a ferrite circulator utilizing the differential phase shift of a transversely magnetized ferrite slab in rectangular waveguide. A gyrator having a differential phase shift of 180° is placed in a closed loop with two magic tees. Consideration of the relative signal phases shows that a signal introduced at A emerges at B, B→C, C→D, and D→A.

Another type of differential phase shift circulator is shown in Figure 28-41. Here, a 90° differential phase element is used together with a short slot
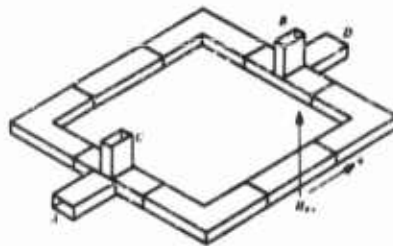
FIGURE 28-40 Transverse field microwave   FIGURE 28-41 Differential phase shift fer-
                    circulator.   rite circulator with schematic symbol.
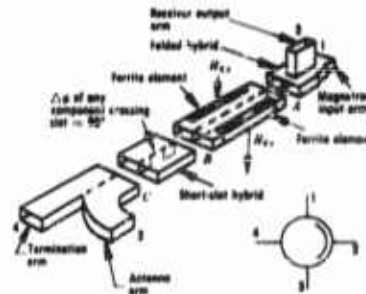
coupler and a folded hybrid tee. This type has an advantage over that shown in Figure 28-40 since only half as much phase shift is required.

Single Sideband Modulator. A single sideband modulator can be built utilizing two quarter-wave plates and a cleverly designed 180° differential phase shifter. The ferrite element is transversely magnetized to give 180° phase differential between signal components parallel to and perpendicular to the magnetic field. The analog of a rotating half-wave plate is created by using two pairs of electromagnets set up perpendicular to each other and to the waveguide axis and excited in time quadrature. By placing the phase shifter between the quarter-wave plates the output wave is continuously shifted in phase, producing single sideband modulation.

Electronic Scanning. The radiation pattern of an antenna can be altered by shifting the phase distribution at the feed. The pattern resulting from waveguide slot antennas can be altered by adding a transversely magnetized ferrite slab in the waveguide to change the relative phases at the slots.

The possibilities of electronic pattern control by the use of ferrites are relatively untapped and seem promising.

28.6.2.2 *Microwave Field-Displacement Applications.* A ferrite slab inserted in the E-plane of a rectangular waveguide can seriously distort the microwave fields. The fields near the ferrite are dependent on the direction of propagation and can be made to differ appreciably for the different directions. Such field displacement effects can be used for a variety of applications.

Field Displacement Isolators. Several methods are used for achieving isolation by microwave field displacement. A common method is to design the structure so that for one direction of propagation the field is displaced into an absorbing load. The absorber may be placed on the ferrite.

Amplitude Modulator. As described before for the longitudinal field case,

the amplitude can be varied by changing the biasing field so as to vary the field distribution in the vicinity of the absorber.

28.6.2.3 Ferromagnetic Resonance Absorption Applications. Resonance absorption is possible regardless of the microwave magnetic field configuration. If a ferrite is located in a linear microwave magnetic field, the resonance frequency is shown by Eq. (14) to be $\omega = |\gamma|(BH)^{\frac{1}{2}}$.

The same devices can be constructed using ferrites with transverse applied magnetic fields as were constructed for ferrites longitudinally magnetized to resonance. A ferrite slab is placed in the E-plane to one side of the center line of a rectangular waveguide and magnetized along the E-plane. Since the ferrite is located in a region of rotating microwave magnetic fields which strongly interact with the precessing electrons for only one direction of propagation, resonance absorption occurs for only the one direction of propagation.

Resonance Absorption Isolator. Resonance isolators constructed in rectangular waveguides with transverse applied magnetic fields give high isolation to insertion-loss ratios and are the most common type of isolator in use. In one of its simpler forms, a ferrite slab is mounted in the guide as shown

in Figure 28-42 and the applied magnetic field is adjusted for resonance. If the microwave magnetic field is rotating in the positive sense with respect to the applied d-c magnetic field, large absorption can be achieved. For the opposite direction of wave propagation, the microwave field rotates in the negative sense in the plane of the ferrite and negligible absorption occurs. A variation of this geometry is to cut

FIGURE 28-42   Transverse field resonance isolator.

out the center section of the ferrite so that two small slabs go only part way across the guide. Front-to-back loss ratios of 100:1 at single frequencies can be obtained using this geometry. In general this geometry has given bandwidths of the order of 10 percent. However, by loading the structure with appropriate dielectric materials, the bandwidth can be increased to approximately 50 percent. Dielectric loading also improves the electrical characteristics by concentrating the microwave energy in the vicinity of the ferrite.

28.6.3 Applications Based on Nonlinear Effects

Probably the most promising area for new applications of ferrimagnetic materials is based on utilizing the nonlinear properties. For small signals the magnetic moments in the ferrite can be considered to precess as a unit in a

uniform precessional mode. However, as the signal level is increased a threshold is reached above which energy is coupled from the uniform precessional mode into higher-order spin-wave modes. This results in anomalies in the propagation conditions which show promise for a variety of applications. The nonlinear effects depend on the type of material, the sample shape, and the manner of loading the microwave structure.

28.6.3.1 Harmonic Generation and Mixing. By driving a ferrite or ferrimagnetic garnet into the nonlinear region, it is possible to produce large second-harmonic signals. Conversion efficiencies of greater than −4 db have been obtained, indicating that ferrites are a practical means of generating high peak powers at high frequencies, when used as frequency doublers. Higher harmonics can be generated but with much lower conversion efficiencies. It also is possible to use the nonlinear properties for mixing of microwaves at high signal levels.

28.6.3.2 Microwave Limiter. The sharp thresholds obtainable for the incidence of nonlinear absorption provide a method for building a passive microwave limiter. Under certain conditions output signals can be maintained substantially below 1 watt for a range of input powers, indicating the possibility of crystal protection by means of a circulator and limiter.

28.7 Limitations

Up to the present time ferrite devices have had certain limitations in frequency coverage, bandwidth, and power-handling capabilities which have restricted their use in ECM. In general the limitations have been due to the properties and geometry of the ferrite and the properties of the guiding structure in which the ferrite is placed. These limitations are now discussed along with some indication of how they are being overcome.

28.7.1 Low-frequency Limitations

The major limitation on the application of ferrites at low microwave frequencies is due to the properties of the ferrite material. In an unsaturated state the domains in a ferrite sample are randomly oriented and shaped. Each domain has a magnetization of $4\pi M_s$ and is acted upon by a field due to its immediate neighbors and by its own demagnetizing field. The neighboring domains and the self-demagnetizing fields can produce any field from zero to $4\pi M_s$. Thus any given domain will exhibit ferrimagnetic resonance at some frequency equal to or less than $4\pi M_s |\gamma|$. The entire sample will exhibit a broad spectrum of losses, called low-field losses, up to a frequency of

$4\pi M_s |\gamma|$. When the sample is saturated* it acts as a single domain and the losses are concentrated about a single frequency given by Eq. (16).

Ferrite devices first came into widespread use at the higher microwave frequencies. These devices for the most part used ferrites operating at or above the resonant frequency rather than below the resonant frequency. This is because of the larger nonreciprocal properties obtained and the smaller applied fields required above the resonant frequency. In order to use the lowest possible applied fields at these high frequencies, longitudinally magnetized rods or slabs magnetized in their plane were used. When these same devices were scaled for operation at lower frequencies a limit was encountered in operating at or above the resonant frequency. Figure 28-31 illustrates the case for a rod. It is seen that $2\pi M_s |\gamma|$ is the lowest frequency at which operation at or above resonance is possible.

One way to obtain operation of ferrite devices at lower frequencies is to reduce the low-field loss by using materials with lower saturation moments. Significant progress has been made in this direction. However, a problem in making materials with low saturation moment is the tendency for the Curie temperature T_c to decrease as it is reduced. It should be recalled that T_c is the temperature at which the ferrimagnetic properties disappear. Another problem resulting from reducing M_s is that the nonreciprocal properties and the resonance absorption are also reduced since both are proportional to M_s. Hence it is necessary to increase proportionately the length of the ferrite and the device when M_s is reduced.

Another approach to low-frequency operation is to use a ferrite configuration, such as a thin disk, which is operable above resonant frequency down to a very low frequency (see Figure 28-31). Although this is effective for some applications, it has the disadvantage of requiring a much larger applied magnetic field. A third possibility is to operate at a frequency below the resonant frequency. At higher frequencies, this type of operation is not advantageous because of the large static magnetic field required and the small interaction which occurs in this region. But as the operating frequency decreases the required field decreases, and the amount of interaction for a low-magnetization material operated above resonant frequency becomes less than that for a high magnetization material operated below resonant frequency. When this happens it is advantageous to operate below the resonant frequency. Under these conditions the ferrite requirements for optimum operation are a high saturation magnetization and a narrow linewidth. The line-

*Saturation occurs when the internal field is sufficient to bring the material above the knee of the magnetization curve. Since most ferrites have narrow hysteresis loops, this internal field at saturation is small. From Eq. (15) it is seen that saturation occurs when the applied field is slightly greater than NM_s.

width is such an important parameter that it is given a separate section below.

In operating at lower and lower frequencies other problems are expected to arise. For example, at lower frequencies one expects other losses such as domain-wall resonances and dielectric relaxation. The present low-frequency limit of devices operating at or above the ferromagnetic resonant frequency is below 2 kmc for some commercially available devices and below 1 kmc in experimental programs. The limit for devices operating below the resonant frequency is not known, but theory indicates it may be under 100 mc.

28.7.2 Linewidth

The ferromagnetic resonance linewidth* is an important parameter in the consideration of limitations of ferrite devices. For almost all applications, the characteristics of devices improve as the linewidth decreases.

The forces existing in a crystal lattice cause the magnetization to have preferred orientations. This anisotropy is described by attributing a fictitious magnetic field to the crystal and calling it the anisotropy field. The linewidth for polycrystalline ferrites is larger than that for single crystals because the anisotropy fields are not all oriented in the same direction. Hence each crystallite is resonant at a different frequency and the absorption line is spread as shown Figure 28-43. Note that off-resonance, the losses for a single crystal sample and a polycrystalline sample are almost the same.
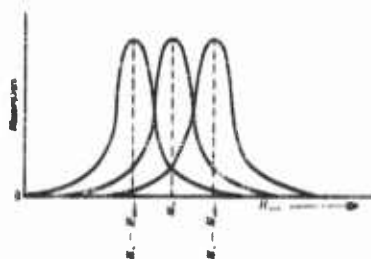
FIGURE 28-43 Ferromagnetic resonance absorption for three crystallites of a polycrystalline sample showing the effect of the different anisotropy field orientation in the sample on the resonant field for each crystallite.

For ferrite devices depending on differential phase shift, it is common to define a figure of merit as the ratio of phase shift to loss. Such a figure of merit is inversely proportional to the resonance linewidth. The derived relationships make it possible to predict the lowest operating frequency for ferrite devices of a given ratio of phase shift to loss. For 90° of differential phase shift, 1 db of insertion loss and a linewidth of 50 oersteds, the theoretical lower frequency limit is approximately 500 mcs. If the linewidth is 5 oersteds, this limit is reduced to 50 mcs.

In order to reduce the linewidth of polycrystalline ferrite samples it is

*The linewidth may be defined as the incremental frequency or field between the points where the absorption in decibels is half the peak resonance absorption in decibels. This number should be independent of the amount of ferrite present.

necessary to reduce the crystalline anisotropy. One way of achieving this is to make small additions of cobalt, since cobalt ferrite has an anisotropy constant opposite to that of other common ferrites. Use of this technique has resulted in polycrystalline ferrites with linewidths of approximately 200 oersteds. To date, the lower limit of linewidth of single crystal ferrites is about 50 oersteds. Polycrystalline ferrimagnetic garnets have linewidths of approximately this value, while single-crystal ferromagnetic garnets have been made with a linewidth of less than 1 oersted.

28.7.8 Bandwidth

There are two important and separate problems associated with the bandwidth of ferrite microwave components. One problem is due to the inherent frequency-response characteristics of the ferrite material. In general, the useful properties of a ferrite result from its tensor permeability; and, as shown in Eq. (4) and (5), the tensor components are frequency dependent. The extent to which this frequency dependence is transferred to a particular device depends on the manner in which the tensor properties are used to produce the device characteristics. A second and often more important problem in ferrite component bandwidth depends on the manner of waveguide loading. The field distribution in the waveguide is frequency dependent. This causes a change in the rf field configuration seen by the ferrite, which in turn may result in a deterioration of the component characteristics.

In order to illustrate some of the factors involved in bandwidth improvement, the specific examples of Faraday rotation devices and resonance-absorption isolators will be considered briefly.

28.7.3.1 Faraday Rotation. As described in Section 28.6, the operation of Faraday-rotation-type devices depends on the difference in permeability for two counterrotating rf magnetic fields. Equation (22) shows that the Faraday rotation per unit length in an infinite medium is frequently dependent. However, if the operating frequency is considerably greater than the ferromagnetic resonance frequency, i.e., $\omega >> |\gamma| H_{eff}$, and if $\omega > 4\pi M_s |\gamma|$, then this frequency dependence is much reduced, as shown by the expansion

$$\frac{\theta_F}{L} = -\frac{4\pi M_s |\gamma| c^{\prime\prime}}{2c}\left[1 + \frac{1}{8}\left(\frac{4\pi M_s |\gamma|}{\omega}\right)^2 \left(1 + \frac{4H_{eff}}{4\pi M_s}\right) + \dots \right] \quad (28\text{-}23)$$

In practice it is usually possible to achieve both of the above conditions; however, this alone does not insure a broadband device. Frequency limitations still arise when one attempts to produce the Faraday rotation in a waveguide having frequency-sensitive propagation characteristics. Faraday

rotators are usually built using a rod of ferrite in a circular waveguide. The guide wavelength has a strong frequency dependence in the dominant-mode operating region. Additional moding and reflection problems can occur because of the high dielectric constant of the ferrite rod. Moreover, if the ferrite rod is too large in diameter, then dielectric waveguide effects will occur for the negative circularly polarized wave at some point in the operating range. These difficulties may be largely overcome by using a small-diameter ferrite rod and a broadband waveguide structure such as a quadruply ridged circular waveguide operated in a region where it has little frequency dependence.

28.7.3.2 Resonance Absorption Isolator. In its simplest form, the resonance isolator consists of a thin slab of ferrite placed in the E-plane of a rectangular waveguide approximately a quarter of the way across the guide. The slab is magnetized transversely, *i.e.*, with the magnetic field perpendicular to the direction of guided wave propagation. There are bandwidth limitations on both the isolation and insertion loss.

The limitation on the isolation bandwidth is that the ferrite must exhibit ferrimagnetic resonance over as broad a band as possible. The use of large linewidth materials and/or the use of nonhomogeneous static magnetizing fields is a simple and adequate solution to this part of the problem.

The limitations on the bandwidth of the insertion is more complicated. If the waveguide fields are h_1^o transverse to the slab and h_2^o along the slab, for circular polarization to exist in the ferrite requires

$$\frac{h_1^o}{h_2^o} = \pm j(\mu \pm \kappa) \qquad (28\text{-}24)$$

It is seen that positive and negative circular polarization will exist at the same point only at the single frequency for which $\kappa = 0$. Therefore, in general, maximum isolation loss and minimum insertion loss occur for different geometries.

In broadband isolator design, the geometry can be chosen to give either minimum insertion loss or maximum isolation loss. The former is used for isolators requiring low insertion loss and moderate isolation. Although isolation is not maximized, it can be made reasonably high. Designing for maximum isolation may be done in narrowband isolators and in broadband units for which minimum length and a maximum isolation are more important than minimum insertion loss.

In the unloaded waveguide the dominant mode fields are characterized by

$$\frac{h_1^o}{h_2^o} = j\sqrt{\left(\frac{f}{f_c}\right)^2 - 1}\ \tan\ (\pi x/a) \qquad (28\text{-}25)$$

Combining Eq. (24), for the condition of low insertion loss, with Eq. (25) yields

$$\mu + K \sim \sqrt{\left(\frac{f}{f_0}\right)^2 - \tan\,(\pi x/a)} \qquad (28\text{-}26)$$

To broad band the low insertion loss of this type of isolator requires a relation of this form to hold over as broad a frequency range as possible. Note that frequency limitations arise from both the characteristics of the waveguide (the right-hand side of the equation) and those of the ferrite (the left-hand side of the equation). The left-hand side of the equation has very little frequency dependence in the region of resonance and, hence, the main bandwidth limitation on the insertion loss is the fact that the polarization of the rf magnetic field is very frequently dependent.

A solution is to use a broadband waveguide structure far above its cutoff frequency so that the ratio of h_1' to h_3' changes slowly with frequency. Several types of broadband waveguide have been used, including double-ridged guide. A particularly useful structure is a waveguide loaded with a dielectric in such a manner as to broad band the waveguide. If the dielectric dimensions are properly chosen, the correct elliptical polarization will be generated in the vicinity of the dielectric and the ferrite can be placed in the region of proper polarization. The dielectric, in addition to keeping the fields relatively fixed in position, serves two other purposes: it concentrates the energy propagating in the waveguide, and it reduces the waveguide impedance. Both of these effects increase the relative magnetic fields, and thus significantly higher isolations result than the same amount of ferrite would produce without the dielectric.

28.7.4 High Power

In general, the limitation on the average power level arises from the fact that losses in the ferrite must appear as heat and the ferrites are poor thermal conductors. Thus the heat is not easily dissipated and, if excessive power is lost in the ferrite, the temperature will increase toward the Curie temperature of the ferrite. The device characteristics deteriorate rapidly as this temperature is approached. The Curie temperature of ferrites now in use ranges from slightly above room temperature to approximately 600°C. Methods of overcoming the high power limitation are to use materials having high Curie temperatures and to provide improved means for cooling the ferrite. Cooling can be accomplished by placing the ferrite against metal walls and by using forced air or liquid cooling.

The power limitation may enter in several ways. In devices such as cir-

culators, gyrators, and off-resonance isolators, in which the ferrite is not at the ferromagnetic resonant frequency, the ferrite dissipates a small fraction of the transmitted power. In some cases the loss due to dielectric and magnetic losses in the ferrite can be made as low as 0.1 db. If it is assumed that the ferrite is able to dissipate 250 watts without overheating and the load VSWR is 2, then the device should be capable of transmitting 10 kw average power.

In resonance absorption isolators the high power limitation is rather severe since both the insertion loss and the isolation loss must be dissipated in the ferrite. Thus the amount of average power which this type of device can handle depends not only on the insertion loss but also on the maximum VSWR of the load. For instance, assuming as above an insertion loss of 0.1 db, a load VSWR of 2 and a maximum dissipation in the ferrite of 250 watts, the isolator is capable of transmitting only 1.9 kw average power.

There also are limitations on the peak power which have not as yet been investigated thoroughly. A major consideration is the nonlinearites of the ferrite properties at very high values of rf field strengths. Since microwave ferrites are good dielectrics, the breakdown limitation normally is not much worse than that of the waveguide structure in which it is placed. However, field distortions can cause localized high electric fields.

Much work has been done to reduce all of the limtations which have been described. Considerable progress has been made, which has increased the usefulness of ferrite devices for ECM. Some examples are given in the following section.

28.8 Applications to ECM

Some simple examples of the use of ferrite devices in ECM are illustrated in Figure 28-44. Part a shows a problem which may arise from attempts to
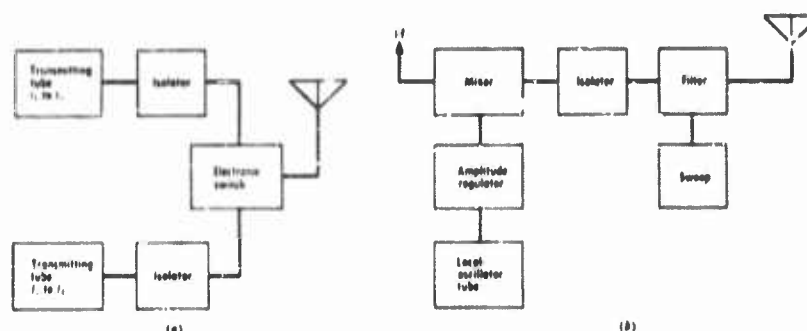
FIGURE 28-44   Typical requirements. (a) Transmitter; (b) receiver.

improve the versatility of an ECM transmitting system. Here it is desired to have either of two transmitters operating into a common antenna. It is assumed that individual transmitting tubes do not have sufficient tuning range to cover the frequency range required. In order to perform the desired function, it is necessary to have a microwave switch. The switch should provide a low attenuation path from the operating tube to the antenna and high attenuation path back into the nonoperating tube. It should be electronically controllable and should be switchable at fairly fast switching rates. If the tube characteristics are sensitive to their loading, isolation of the tubes from the rest of the circuit also may be required. The ideal isolator has low attenuation, or insertion loss, for the forward direction and large attenuation, for isolation, for the reverse direction. The isolator must be capable of handling the output power of the transmitting tubes and must have a bandwidth at least equal to the frequency range of operation, or be tunable over this range.

Figure 28-44b shows an example of a countermeasures receiver application. It is assumed that the local oscillator output has undesirable amplitude variations. The amplitude regulator should reduce these variations to a small value over the entire frequency range of operation so that the input to the mixer remains relatively constant. It may be necessary to have a tunable filter to avoid the ambiguity of image signals. A bandpass filter with high attenuation at the image frequency is preferred, but a band rejection filter to eliminate the image might be utilized. An isolator also may be needed to prevent local oscillator radiation which might reveal the location of the receiver to enemy forces. Such an isolator would require low attenuation for incoming signals. A bandwidth of up to an octave, or even greater, may be required.

Figure 28-45 illustrates schematically another systems problem in which there are many transmitters, each driving one element of an antenna array. It is desired to provide controllable phase shift in the various antenna elements in order to control the antenna pattern.

All of the microwave control device problems illustrated above can be approached by utilizing the properties of special materials. In some cases, this is the only approach to achieving a solution. Some examples of specific results will now be given.

28.8.1 Broadband Low Power Isolators

One of the major advances in adapting ferrite devices to the needs of ECM has been the development of techniques for broadband ferrite isolators. It is now possible to get isolators covering full waveguide bandwidths, approximately 50 per cent, from S-band through K-band. Isolators having approximately octave band coverage are available from L-band to X-band.
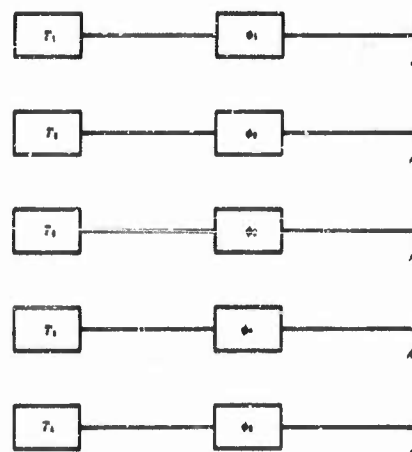
Figure 28-45 Systems requirement for electronically controllable microwave phase shifters. T = transmitter. ϕ = electronically controlled phase shifter. A = element of antenna array.

These advances have been made from a "state of the art" of approximately 10 per cent bandwidth in ferrite devices in less than four years. Some of the techniques developed can be applied also to other devices.

Some typical examples of broadband waveguide isolator characteristics for G, J, and X-band are shown in Figures 28-46, 28-47, and 28-48. Characteristics of coaxial line isolators in L and S band are shown in Figures 28-49 and 28-50. These data are all for dielectric loaded resonance type isolators.

28.8.2 Broadband High Power Isolators

Considerations of heat dissipation generally predominate in the design of high power isolators. These conditions usually require configuration different from those giving optimum performance at low power. Nevertheless some of the techniques used in broadbanding low power isolators also can be utilized at high power.

The performance characteristics of a broadband high power resonance isolator are shown in Fig. 28-51. At milliwatt input power, the isolation exceeds 20 db over most of the 800 mc bandwidth falling to 19 db at the low frequency band-edge. For a simulated input power of 3200 watts and load VSWR of 2, i.e., 800 watts absorbed, the isolation is still greater than 15 db except for the one point at band edge. The full capacity of this type of isolator exceeds the values shown.
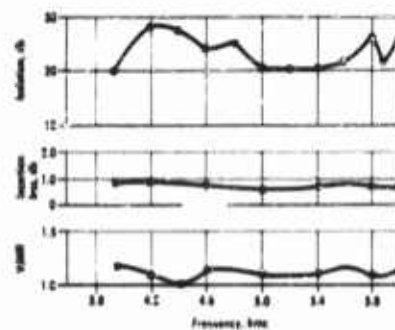
FIGURE 28-46 Broadband waveguide isolator characteristics, G-band.
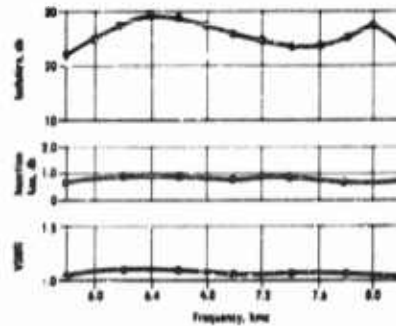
FIGURE 28-47 Broadband waveguide isolator characteristics, J-band.
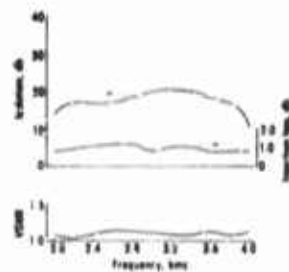
FIGURE 28-48 Broadband waveguide isolator characteristics, X-band.
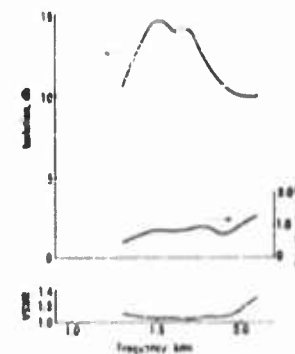
FIGURE 28-49 Broadband coaxial line isolator characteristics, 1.3-2.1 kmc.
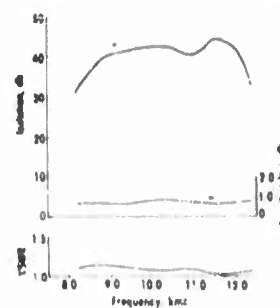
FIGURE 28-50 Broadband coaxial line isolator characteristics, 2-4 kmc.
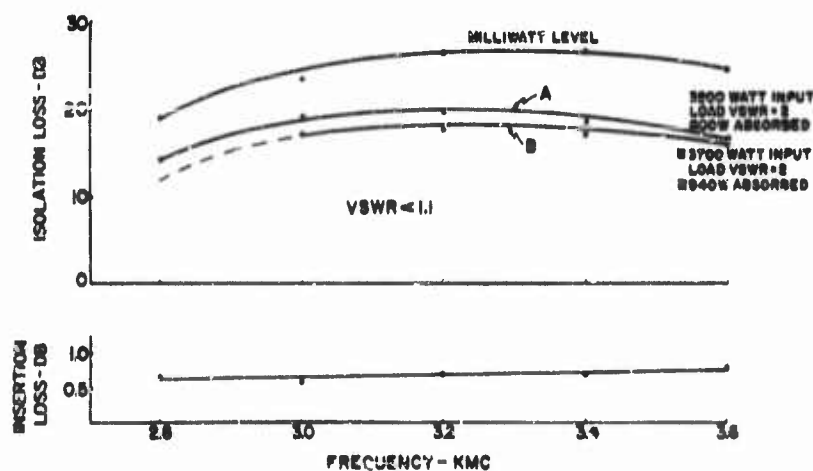
FIGURE 28-51 Performance data for a high-power load isolator using $Ni_{0.7}Cu_{0.8}Mn_{0.08}$ $Fe_{1.9}O_4$ and nickel ferrite, Ferroxcube 106.

28.8.3 Broadband Amplitude Regulator

An electronically controllable microwave attenuator utilizing the "low field" magnetic loss characteristics of ferrites has been developed to regulate the output of microwave source to a small variation over a 2-to-1 frequency range. Figure 28-52 shows typical results with the ferrite attenuator placed be-
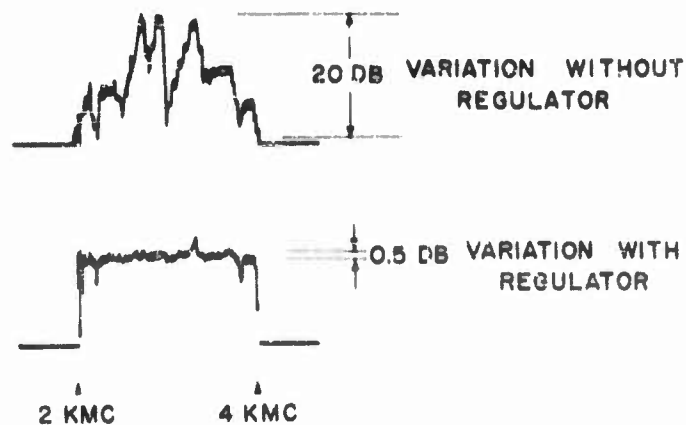
FIGURE 28-52 Regulation of mixer crystal current.

tween the local oscillator and the mixer. The oscillator was the French CSF CO-127, and the mixer was a broadband coaxial mixer. The upper curve shows the crystal current variation with a fixed 20 db pad used between the local oscillator and the mixer. Variations close to 20 db are indicated. The lower curve shows the mixer current variation remaining when the fixed pad was replaced by the ferrite attenuator, and the feedback loop was closed. The deviations from the relatively flat response occurred when input variations exceed the dynamic range of the attenuator. The particular ferrite regulator used was designed for a 15 db dynamic range.

The attenuator consists of a ferrite sleeve mounted in a coaxial line and transversely magnetized by a small electromagnet. Its size is less than 2 inches cubed; it weighs less than 12 ounces.

28.8.4 Broadband Microwave Switch

Various types of microwave switches have been developed using ferrites. One type that is capable of giving broadband operation utilizes two broadband coaxial line couplers and a broadband switchable ferrite phase shifter. Developmental models of such a switch have been operated over ½ octave bandwidths. In principle, it should be possible to get full octave frequency coverage.

28.8.5 Broadband Controllable Phase Shifters

The same techniques used to give broadband switchable phase shift can be used to obtain continuously variable phase shift. Developmental models have been operated over greater than ½ octave with figures of merit exceeding 150 degrees per db loss.

28.8.6 General Comments

The above examples illustrate only a few of the many possible applications of ferrites to ECM. Although in principle all of the devices mentioned in Section IID could be used in ECM, the actual use so far has not been great. This has been due in part to the limitations discussed in the Section IIE. The newness of the entire field of ferrite devices leaves many important devices still in developmental stages. Another contributing factor is that systems engineers often are not cognizant of the newer developments or may be unwilling to use them until their utility is proved. However, with improving technology and increasing awareness by systems engineers of the potential in this field, the future uses of ferrites in ECM should increase significantly.

III. Gaseous Electronics

28.9 Fundamental Characteristics

Useful microwave characteristics can be obtained from interactions of

microwave signals with free electrons. In the absence of collisions these interactions are entirely reactive, with the electron motion lagging the microwave electric field by 90°. In the presence of a gas, in which collisions between electrons and neutral gas atoms are assumed to act as a continuous viscous damping force, Newton's second law of motion may be written

$$m \frac{dv}{dt} + rv = - eE_0 \exp j\omega t = - Ee \qquad (28\text{-}27)$$

where m = mass of the electron

v = electron velocity

r = damping parameter due to collisions

$-e$ = charge on electron

ω = angular frequency of the microwave signal

$E = E_0 \exp j\omega t$ = microwave electric field

Assuming the damping parameter r is a constant independent of velocity, the electron drift velocity determined from Eq. 27 is

$$v = eE/(r + j\omega m) \qquad (28\text{-}28)$$

The current density is drift velocity multiplied by the electron number density n and the charge on a single electron.

$$J = - ncv = ne^2E/(r + j\omega m) \qquad (28\text{-}29)$$

The complex conductivity $\sigma_c = \sigma_r + j \sigma_i$ is

$$\sigma_c = J/E = ne^2/(r + j\omega m) \qquad (28\text{-}30)$$

Note that the conductivity is directly proportional to the electron density. Equation 30 has the same form as the admittance of a resistive-inductive (RL) series circuit. It is well known that maximum power is transferred to such a circuit when $R = \omega L$. Similarly, maximum power is transferred to the electrons when $r = \omega m$. If $r = \omega v_c$, v_c is the collision frequency of electrons and gas atoms, the maximum power is transferred when $\omega = v_c$. For helium, v_c is approximately 2.3×10^9 where p is the pressure in millimeter of mercury.

Any useful device based on interactions of electrons and microwave fields must provide a method of producing the necessary electrons and must give consideration to electron loss mechanisms. A rigorous analysis would have to consider both the particle distribution and the energy distribution.

28.9.1 Electron Loss Mechanisms

The major mechanisms for loss of electrons are diffusion, attachment and recombination.

Diffusion is the random scatter of particles away from a region of concentration. The characteristic diffusion equation is obtained by combining an equation relating the current density to the particle density gradient with the equation of particle continuity. Diffusion losses predominate over a wide range of practical conditions. They normally are characterized by a parameter Λ, called the characteristic diffusion length, which is a measure of the size of the container. When the density is high the electrons and ions will diffuse with essentially the same velocity, a condition known as ambipolar diffusion.

The phenomenon of electron attachment to a neutral atom is a common occurrence for gases whose outer electronic shells are nearly filled. The measure of the ease with which an electron can attach to a neutral atom or molecule is given by the electron affinity energy. For those gases which exhibit attachment this varies from about 4 volts (for gases like F and O_2) to nearly zero. It is negative for those which do not inert to extra-atomic electrons. These atoms, which have $'S_0$ ground states, include the noble gases. Since noble gases would normally be used in building a device, the effects of attachment can often be neglected. However, electron attachment to impurity atoms can be of great practical importance. Also of practical importance is the use of attaching gases to speed up the removal of electrons in some applications, such as TR tubes or other microwave switches.

Another common loss mechanism is the recombination of negative ions and electrons with positive ions to form neutral atoms. If the positive and negative particle concentrations are equal, which is true in almost all gas discharges, this mechanism exhibits a linear relationship between the reciprocal of partical density and time. It can predominate under high pressure and high electron density conditions. If true electron ion recombination is occurring, the recombination coefficient characterizing the process should be essentially independent of pressure.

28.9.2 Microwave Breakdown

In a high frequency gas discharge breakdown, the primary ionization due to the electron motion is the only production phenomenon which controls the breakdown, and for this reason it is the simplest type to consider. The calculated value for the maximum kinetic energy in the oscillatory motion of an electron at the minimum field intensities for breakdown (experimentally determined) corresponds to about 10^{-8} volt. It is therefore obvious that the energy of oscillation is insufficient to account for breakdown.

As noted above, a free electron in a vacuum under the action of an alternating electric field oscillates with its velocity 90° out of phase with the field, and thus takes no power, on the average, from the applied field. The electron can gain energy from the field only by suffering collisions with the gas atoms, and it does so by having its ordered oscillatory motion changed to random motion on collision. The electron gains random energy on each collision until it is able to make an inelastic or ionizing collision with a gas atom. The fact that the electron can continue to gain energy in the field, on the average, despite the fact that it may either move with or against the field can be seen by noting that the energy absorbed is proportional to the square of the electric field and hence is independent of its sign.

A gas subjected to high frequency electric fields will break down and become conducting when the number of electrons produced per second becomes equal to or greater than the number of electrons lost per second. The production mechanism in a high frequency discharge is ionization within the body of the gas. The important loss mechanisms are diffusion, attachment and recombination as discussed above, with diffusion predominating under a wide variety of experimental conditions. However, diffusion theory will not apply where the electron mean free path, the average distance between collisions, is comparable to the dimensions of the container. Another limit on the diffusion mechanism occurs when the electron oscillation amplitude becomes so great that the electrons would hit the walls on each cycle.

Other methods of providing a source of electrons include thermionic emission and gas discharges produced by direct current.

28.9.3 Propagation Characteristics in the Presence of an Applied Magnetic Field

Free electrons in a magnetic field constitute a dispersive medium. Electrons are constrained to move in spirals around the magnetic field lines at the cyclotron frequency $\omega_c = eB/mc$. If a high frequency electric field perpendicular to the magnetic field oscillates at the cyclotron frequency, the electrons move in phase with the electric field and they gain energy from it continuously. When the frequency of the impressed field is different from the cyclotron frequency the electron motion is alternately in and out of phase with the field so the electrons alternately gain and lose energy; the electrons have a large reactive interaction with the field but gain no energy from it. The dielectric constant of the medium depends on the electron density and on the relative frequency of the propagating signal and the cyclotron orbital motion. As the dielectric constant varies, the wavelength in the medium varies and hence also the phase of the signal that has passed through the medium.

When the electrons are obtained by ionizing a gas they have collisions with electrons and other particles which tend to disorder their motion. Each collision represents an exchange of energy which results in an attenuation of the microwave signal propagating through the medium. At modest electron densities most of the collisions take place between electrons and neutral molecules at a rate determined by the collision frequency ν_c, which in turn depends largely on the pressure and type of gas. Hence the collision frequency and indirectly the gas pressure and type of gas determine the attenuation of the propagating signal.

The presence of the applied static magnetic field to a gas plasma creates an electrical anisotropy which manifests itself in the current vector. Under small signal conditions and a plasma containing a small percentage ionization this may be represented by

$$\mathbf{J} = |\sigma| \cdot \mathbf{E} \qquad (28\text{-}31)$$

where

$$|\sigma| = \begin{pmatrix} \sigma & -jK & 0 \\ jK & \sigma & 0 \\ 0 & 0 & \sigma_p \end{pmatrix} \qquad (28\text{-}32)$$

Neglecting loss, *i.e.*, assuming $\nu_c = 0$,

$$\sigma = j(ne^2/mc)\omega/(\omega_c^2 - \omega^2) \qquad (28\text{-}33)$$

$$K = -j(ne^2/mc)\omega_c/\omega_c^2 - \omega^2) \qquad (28\text{-}34)$$

Please note that these tensor elements have the same mathematical form, with the same kind of dependence on frequency, as those of the susceptability tensor for a ferrite given in Equations 4 and 5 of Section IIA. This is seen more clearly if one relates ω_c to γH_0 and compares σ and K with κ and χ respectively.

The propagation equation is

$$\nabla \times \nabla \times \mathbf{E} - \beta_0^2 |\epsilon| \mathbf{E} = 0 \qquad (28\text{-}35)$$

where

$$|\epsilon| = \epsilon_0(I - j\sigma/\omega\epsilon_0) \text{ and } I = \text{unit diagonal tensor} \qquad (28\text{-}36)$$

Note that this is the same mathematical form in E and the dielectric tensor

ϵ as for h and the permeability tensor μ in Eq. 8 of Section IIA. Therefore the formal solution for the propagation constant also is the same. This demonstrates a complete analogy between propagation in anisotropic gas plasmas and propagation in a ferrite media.

A simplified comparison of these media is made in Figure 28-53. Both of

FIGURE 28-53   Comparison of dispersive media.

these media derive their useful microwave properties from interactions of the microwave signals with electrons in the media. In the case of the ferrite these interactions are with precessing bound electrons, while in the case of the gas discharge the interactions are with planetary free electrons. In both cases, the natural frequency of electron motion is directly proportional to an effective magnetic field, and in simplified cases the constant of proportionality is exactly the same. The dispersive parameter for the ferrite is the permeability and for the gas discharge the conductivity or dielectric constant. In the simplest cases, the tensor permeability and tensor conductivity have the same tensor form. From the application standpoint, it is important to recall that in the case of the ferrite media the coupling is to the rf magnetic field and in the case of the gas discharge the coupling is to the rf electric field. Since the natural frequency of motion is circular, the interaction in both cases will be greatest if the frequency of the microwave signal coincides with the natural resonance frequency and if the microwave signal is circularly polarized in the same direction as the electron motion. In both cases, large resonance absorption will occur when the frequencies coincide and large dispersive effects will occur in the vicinity of resonance.

28.10 Microwave Applications

The use of ionized gases in the control of guided microwaves is practically as old as guided microwaves. The most obvious example is the tube known as a transmit receive tube or TR tube which is in use in radar systems. In such a tube the incoming high power rf signal causes the gas in the tube to break down across a gap and change the impedance of the gap in such a way

as to protect some component on the other side of the tube from the high power signal which causes the breakdown. Gas discharges have been used as noise sources at microwave frequencies. These better known applications of gas discharges will not be discussed here.

The properties of gas plasmas under the influence of an external applied magnetic field are particularly interesting for microwave applications. By utilizing the interaction of electrons in cyclotron orbits with microwave fields oscillating at or near cyclotron resonance, wide ranges of reactive or resistive impedances can be obtained. The impedance changes depend on the type of gas used and on such parameters as electron density, electron collision frequency (related to gas pressure), and difference between the signal frequency and electron cyclotron frequency. Since the propagation constant has the same form as for the ferrite, the same application principles apply and in principle all the types of devices described in Section 28.6 should be achievable by gaseous electronics techniques.

28.10.1 Faraday Rotation

By applying a longitudinal magnetic field to a gas plasma confined in a circular waveguide, rotation of a linearly polarized microwave signal can be produced just as in the case of a ferrite. Figure 28-54 illustrates the magneto-
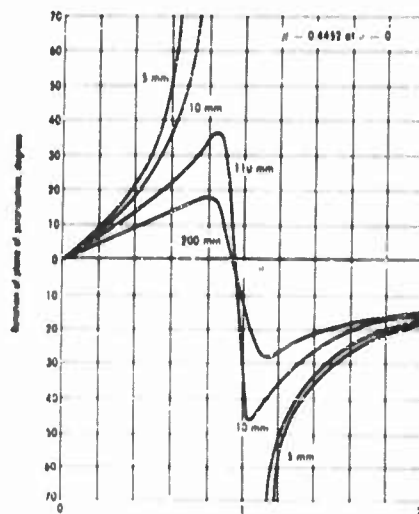
Figure 28-54 Rotation of the polarization plane of the transmitted TP wave vs the magnetic field variable σ for several values of gas pressure.

rotation of TE_{11} linear polarized waves through a neon plasma as a function of magnetic field for various gas pressures. It is apparent that an increase in the electron collision frequency (corresponding to an increase in gas pressure) reduces the amount of rotation obtainable with a given value of magnetic field. However, for the smaller pressures sufficient rotation is obtained to be usable for microwave applications. That this has not been utilized to a greater extent the case is due largely to the practical problems of achieving stable operation and adequate life.

28.10.2 High Power, Broadband Microwave Gas Discharge Switch

An application currently under development specifically to meet ECM requirements is directed toward achieving a high power microwave switch by utilizing magnetic field control of the gas breakdown conditions. In its simplest form, the switch tube consists of a section of S-band rectangular waveguide sealed off at both ends with vacuum windows and containing a rare gas or mixture of rare gases at a pressure of the order of a tenth of a millimeter of mercury.

The tube is 'fired' by the incident microwave power, in a time of the order of one hundredth of a microsecond, when the proper magnetic field is applied to it. The resultant high electron density of the plasma acts like a short circuit at the inside surface of the input window. When the applied magnetic field is zero or sufficiently small, the tube is 'unfired' in the presence of the incident microwaves and acts like a low-loss section of waveguide. Thus, the 'on-off' action of the tube can be controlled by varying the magnitude of the applied magnetic field.

The tube's operation is based on the phenomenon of electron cyclotron resonance. An electron in a static magnetic field spirals around the flux lines at the electron cyclotron frequency. When a magnetic field is applied in a direction perpendicular to the electric field in the switch tube, two principle effects occur. First, the energy transferred from the electric field to an electron exhibits a sharp resonance when the cyclotron frequency coincides with the angular frequency of the microwaves. The magnetic field for which this occurs is called the cyclotron resonant field or B_r and has a value of around 1200 gauss at S-band. Secondly, the electron's diffusive motion transverse to the magnetic field is retarded. These two effects determine the breakdown power vs. magnetic field characteristics upon which the design of the switch tube depends.

Figure 28-55 shows a typical curve of breakdown power in watts peak as a function of magnetic field. The magnetic field parameter in the graph is normalized to unity at the resonance condition. If the magnetic field is not adjusted for resonance, the motion of the electrons is out of phase with the
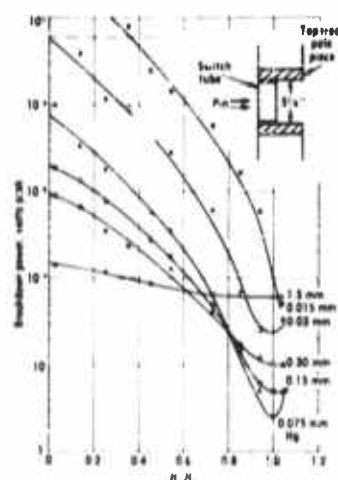
FIGURE 28-55 Microwave breakdown power vs magnetic field in argon. Tube length: ¼ inch. Pulse width, 2 microseconds. PRF: 500 per second. Frequency: 2.84 kmc. B-field: transverse.

microwave electric field during some portions of their orbits. Under resonance conditions, electrons rotate around the magnetic flux lines at the same frequency as the microwave electric field and can gain energy continuously from it. Hence, at cyclotron resonance there is a maximum transfer of energy from the microwave electric field to the electrons. At sufficiently low pressure, the energy transfer at cyclotron resonance can be of the order of 10^4 or more times that at zero magnetic field.

The switch tube is designed so that with zero applied magnetic field, the power required for the gas to break down is larger than the source power. Under these conditions an incident pulse does not produce a discharge and the switch is 'open' or 'unfired'. When a magnetic field near cyclotron resonance is applied, the power required for breakdown is much less than the source power and an incident pulse produces a strong discharge. The resultant high density discharge acts like a short circuit and the switch is 'closed' or 'fired'.

Figure 28-56a illustrates how this principle may be used to alternately switch two transmitters to a common antenna. With transmitter No. 1 in operation, it is desired to have transmitter No. 2 completely isolated from the circuit. Hence, the switch in Arm 1 should be unfired by having a magnetic field applied which is removed from resonance and the switch in Arm 2 fired by biasing to resonance. When transmitter No. 2 is in operation, the reverse is required. A switch of this type has been built. Other possible configurations are illustrated in Fig. 28-56b, c, and d. These configurations permit the simultaneous operation of both transmitters with the power from the unwanted transmitter routed into a dummy load. For configurations b and c, with the switches unfired, the unused transmitter is isolated from the antenna by only the attenuation of the coupler. The configuration shown in d permits switching the unused transmitter into a dummy load and, in addition, isolated the unused transmitter from the antenna by the full attenuation of the gas switch.
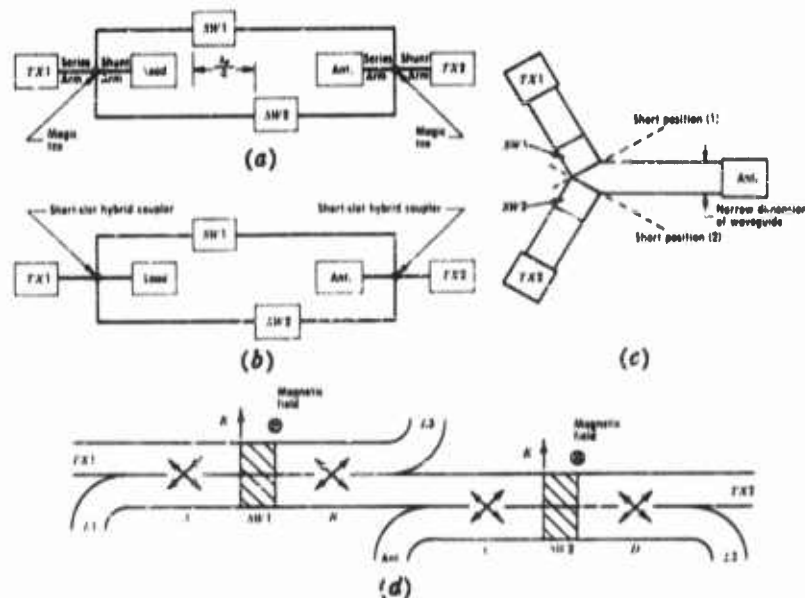
FIGURE 28-56 Possible switch configurations (a) Balanced magic tee (or ring hybrid) switch. (b) Balanced short-slot hybrid switch. (c) 120° E-plane "Y" switch. (d) Cyclotron-resonant gas switch configuration. SW1 and SW2 are switch tube pairs. A, B, C, and D are 3-db top-wall short-slot hybrids. $L1$ and $L2$ are high-power terminations. $L3$ is a low-power termination.

$$\text{Switch State I} \quad \begin{Bmatrix} TX1 \rightarrow ANT \\ TX2 \rightarrow L2 \end{Bmatrix} \longleftrightarrow \begin{Bmatrix} SW1 \text{ Unfired} \\ SW2 \text{ Fired} \end{Bmatrix}$$

$$\text{Switch State II} \quad \begin{Bmatrix} TX1 \rightarrow L1 \\ TX2 \rightarrow ANT \end{Bmatrix} \longleftrightarrow \begin{Bmatrix} SW1 \text{ Fired} \\ SW2 \text{ Unfired} \end{Bmatrix}$$

28.10.3 Microwave Phase Shift

Preliminary investigations have been made to determine the microwave phase shift that can be produced by means of an ionized gaseous plasma. One approach is to use the dispersive properties of a gas near cyclotron resonance. In principle, an electronically controllable phase shifter can be made by producing free electrons in some structure suitable for propagating a microwave signal whose electric field is perpendicular to an applied magnetic field. The dielectric constant of the medium depends on the electron density and on the relative frequencies of the propagating signal and the cyclotron orbital motion. As the dielectric constant varies, the wavelength in the

medium varies and hence also the phase of the signal passing through the medium.

Coaxial geometry was chosen for this study because of the wide range of frequencies it will transmit, and because, with two separate conductors that can be used as electrodes, it also offers more possibilities for creating the electrical discharge. The basic configuration is shown in Figure 28-57. The gas
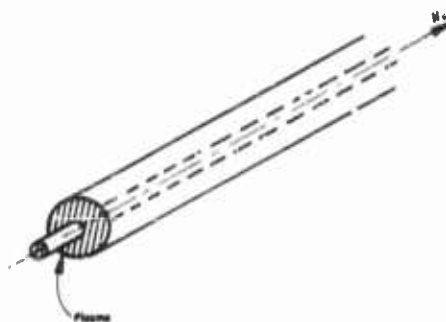
FIGURE 28-57 Gas discharge phase-shift configuration.
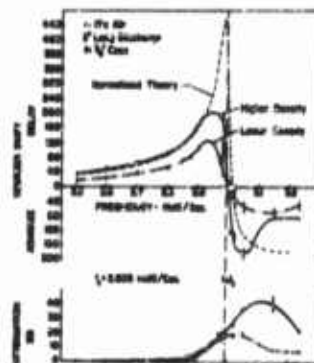
FIGURE 28-58 Microwave phase shift utilizing electron cyclotron resonance.

plasma is confined between the inner and outer conductors of a section of line and the magnetic field as applied longitudinally.

Figure 28-58 shows some preliminary experimental measurements of the microwave dispersion in S-band for a microwave discharge in air. The results demonstrate that large phase shifts can be obtained when the microwave actual frequency is slightly above or below the frequency for electron-cyclotron resonance. It is seen that the insertion loss is high in the immediate vicinity of resonance as would be expected, but the phase shift remains large for frequency regions where the insertion loss is low. The theoretical predictions normalized to the experimental data at the frequency 2.8 kmc, are superimposed for comparison. The dispersion curves observed are in good agreement with the theoretical predictions, except in the immediate vicinity of resonance where it is difficult to make accurate comparison.

Figure 28-59 illustrates how phase shift and attenuation vary with applied magnetic field. It shows a change from 200 to 300 degrees phase shift for a change from about 50 gauss. This indicates that a narrow band device could be built to use magnetic field tuning over a range of 100 degrees of phase shift with a nearly linear tuning rate of about 2 degrees per gauss. Figure 28-60 is taken for a discharge in helium at fixed magnetic field for various

FIGURE 28-59 Gas discharge phase shift and attenuation as a function of magnetic field.

FIGURE 28-60 Microwave phase shift utilizing electron cyclotron resonance (discharge in helium).

conditions of electron density, i.e., higher discharge power corresponding to higher density. The data illustrates that control of the density is another effective way to vary the microwave phase shift. Note that approximately 250° change in phase shift was obtained in certain regions of the curves.

28.11 Applications to Operational Problems

In addition to providing a basis for microwave control devices, a knowledge of the properties and propagation characteristics of gas discharges can help to solve certain operational problems. One such problem, illustrated in Fig. 28-61, has to do with the effect of a microwave electrical breakdown near antennas on high flying vehicles. Antennas on such vehicles may initiate electrical discharges because the electric field required to break down the air at low pressures is in general much less than that required at atmospheric pressure. Considerable information pertinent to this problem can be obtained from a knowledge of the propagation and breakdown conditions in air.

28.11.1 Microwave Breakdown at High Altitudes

In order to provide information pertinent to this problem, calculations have been made of threshold electric fields for cw breakdown and maximum electric field for pulse transmission at various frequencies from 100 mc/sec to 35 kmc/sec at altitudes up to 500,000 feet. In order to perform such cal-

FIGURE 28-59 Gas discharge phase shift and attenuation as a function of magnetic field.

FIGURE 28-60 Microwave phase shift utilizing electron cyclotron resonance (discharge in helium).

conditions of electron density, i.e., higher discharge power corresponding to higher density. The data illustrates that control of the density is another effective way to vary the microwave phase shift. Note that approximately 250° change in phase shift was obtained in certain regions of the curves.

28.11 Applications to Operational Problems

In addition to providing a basis for microwave control devices, a knowledge of the properties and propagation characteristics of gas discharges can help to solve certain operational problems. One such problem, illustrated in Fig. 28-61, has to do with the effect of a microwave electrical breakdown near antennas on high flying vehicles. Antennas on such vehicles may initiate electrical discharges because the electric field required to break down the air at low pressures is in general much less than that required at atmospheric pressure. Considerable information pertinent to this problem can be obtained from a knowledge of the propagation and breakdown conditions in air.

28.11.1 Microwave Breakdown at High Altitudes

In order to provide information pertinent to this problem, calculations have been made of threshold electric fields for cw breakdown and maximum electric field for pulse transmission at various frequencies from 100 mc/sec to 35 kmc/sec at altitudes up to 500,000 feet. In order to perform such cal-

FIGURE 28-61 Schematic representation of operational problems.

culations, it is, of course, necessary to consider the composition of the at-
mosphere and the role of the various fundamental processes contributing to
break-down.

For preliminary calculations, the assumption can be made that the com-
position of air does not change, insofar as it affects electrical breakdown at
high frequencies, over the entire altitude range under consideration. Although
involving some error, this is a reasonable first approximation and probably
leads to results on the conservative side.

With regard to mechanisms, a gas subjected to high frequency electric
fields will break down and become conducting when the number of electrons
produced per second becomes equal to or greater than the number lost per
second. The production mechanism in a high frequency discharge is ionization
within the body of the gas. The important loss mechanisms in air are diffusion
and attachment, with attachment predominating in an unconfined region.
Recombination could be significant at high electron concentrations, but it is
probably at least an order of magnitude less important than attachment in
determining breakdown at high altitudes. Accordingly, only attachment need
be considered in preliminary calculations. To the extent that the other pro-
cesses actually contribute to the loss of electrons, the resultant conclusions
will be on the conservative side.

Although calculations for cw conditions are interesting, the case of greater
pertinence to the practical problem is pulsed transmission in which some in-

crease in electron concentration is allowable. It is well known that signals may be transmitted through appreciable concentrations of electrons, as illustrated by the gas discharge phase shift studies described previously, and so build up of electrons in front of the antenna may be tolerated provided the electron density is not too great. The electron concentration above which there is effectively no transmission is the so-called "plasma resonant density" which is equal approximately to 10^{18} divided by the square of the wavelength in centimeters. For a concentration equal to half plasma resonant density there will be almost complete transmission. Accordingly, the plasma resonant density may be considered to be a fairly sharp upper limit.

For a pulse in which the electric field is larger than that required for breakdown, the electron concentration will start to increase at an exponential rate determined by the difference in the ionization rate and the electron loss rate. If the concentration at the end of the pulse just reaches the plasma resonant density, practically the entire pulse will have been transmitted. Accordingly, calculations can be made to predict when this should occur. An example of the kind of data obtained is shown in Figure 28-62 for the frequen-
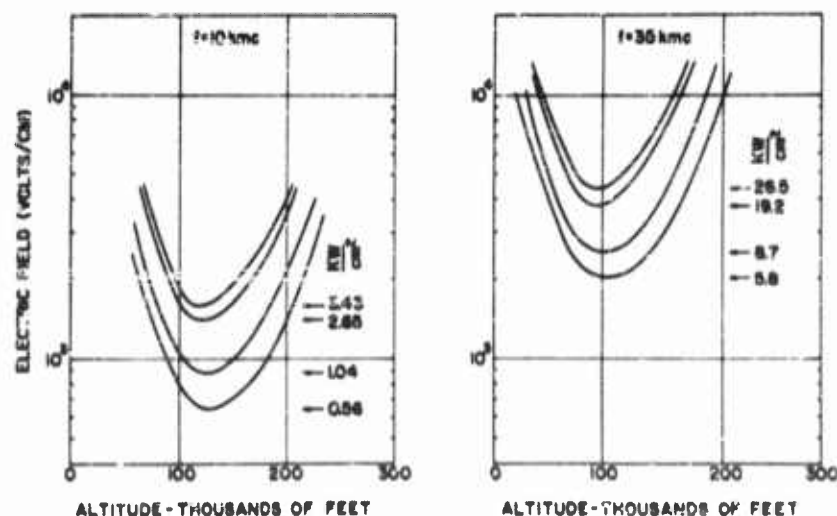
FIGURE 28-62　Peak microwave electric field for pulse transmission vs altitude (for different pulse lengths at frequencies of 10 and 35 kilomegacycles

cies of 10 kmc/sec and 35 kmc/sec, and for pulse lengths ranging from .05 to 5 microseconds. Under the conditions assumed, it is believed that at least 90% of the pulse will be transmitted at the microwave electric fields shown

on the curves. Also shown is the power radiated from a linear array in kilowatts per square centimeter for conditions corresponding to the minimum of each curve.

The calculations indicate quite clearly that the higher frequencies allow the transmission of considerably greater amounts of power than do the lower frequencies and, as expected, that the threshold field allowable for breakdown is lower for long pulse lengths than for short pulse lengths.

TABLE 28-1. TUNING CHARACTERISTICS OF RESONANT LINE STRUCTURE
Resonant Line

Capacitance ($\mu\mu f$)	Unsaturated		Saturated	
	Q	F (mc)	Q	F (mc)
15	*	89	30	175
20	*	82	50	152
25	*	76	90	137
30	*	70	95	126
40	12	64	110	110
50	30	58	122	100
60	75	54	130	93
70	85	51	140	87
			F_{max}/F_{min}	1.9

*Too low to read accurately.

BIBLIOGRAPHY

1. Snoek, J. L., "Nonmetallic Magnetic Material for High Frequencies," *Philips Tech. Rev.*, Vol. 8, pp. 353-360, 1946; *New Developments in Ferromagnetic Materials*, Elsevier Publishing Co., Amsterdam, N. Y., 1947.

2. Neel, L., "Propertiés Magnetiques des Ferrites; Ferrimagnetism et Antiferromagnetisme," *Ann. Phys.*, Vol. 3, pp. 137-198, 1948.

3. Kittel, C., "On the Theory of Ferromagnetic Resonance Absorption," *Phys. Rev.*, Vol. 73, pp. 155-161, 1948.

4. Kittel, C., "Physical Theory of Ferromagnetic Domains," *Revs. Modern Phys.*, Vol. 21, p. 1, 1949.

5. Rado, G. T., "Ferromagnetic Phenomena at Microwave Frequencies," *Advances in Electronics*, Vol. 2, p. 251, 1950.

6. Bozorth, R. M., *Ferromagnetism*, Van Nostrand, N. Y., 1951

7. Hogan, C. L., "The Ferromagnetic Faraday Effect at Microwave Frequencies and Its Applications. The Microwave Gyrator," *Bell System Tech. J.*, Vol. 31, pp. 1-31, 1952.

8. Fairweather, A., F. F. Roberts, and A. J. E. Welch, "Ferrites," *Repts. Progr. Phys.*, Vol. 15, p. 142, 1952.

9. Polder, D., and J. Smit, "Resonance Phenomena in Ferrites," *Revs. Modern Phys.*, Vol. 25, p. 89, 1953.

10. Rowen, J. H., "Ferrites in Microwave Applications," *Bell System Tech.*, Vol. 32, p. 1333, 1953.

11. Suhl, H., and L. R. Walker, "Topics in Guided Wave Propagation Through Gyromagnetic Media," *Bell System Tech. J.*, Vol. 33, pp. 579-659, 939-986, 1133-1194, 1954.

12. Lax, B., K. J. Button, and L. M. Roth, "Ferrite Phase Shifters in Rectangular Waveguide," *Tech. Memo No.* 48, MIT Lincoln Laboratory, Nov. 2, 1953; *J. Appl. Phys.*, Vol. 26, pp. 1413-1421, 1954.

13. Fox, A. G., S. E. Miller, and M. T. Weiss, "Behavior and Applications of Ferrites in the Microwave Region," *Bell System Tech. J.*, Vol. 34, pp. 5-103, 1955.

14. Aden, A. L., W. P. Ayres, J. L. Melchor, and P. H. Vartanian, "Ferrites—Part I, Fundamentals," *The Sylvania Technologist*, Vol. VIII, No. 3, 1955; "Part II, Microwave Applications," *The Sylvania Technologist*, Vol. VIII, No. 4, 1955; "Part III, Limitations and Measurements," *The Sylvania Technologist*, Vol. IX, No. 1, 1956.

15. *Proc. IRE*, Ferrite Issue, Vol. 44, No. 10, pp. 1229-1516, October 1956.

16. *IRE Transactions on Microwave Theory and Techniques*, PGMTT National Symposium Issues, Vol. MTT-6, No. 1, January 1958; Vol. MTT-7, No. 1, January 1959; Vol. MTT-8, No. 1, January 1960.

17. Butler, T. W., W. J. Lindsay, and L. W. Orr, "The Application of Dielectric Tuning to Panoramic Receiver Design," *Proc. IRE*, Vol. 43, September 1955.

18. Butler, T. W., H. Diamond, and L. W. Orr, "Subminiature Nonlinear Capacities for Application to VHF, Wide-Range Tuning Devices," *Proc. NEC*, Vol. XI, 1955.

19. Butler, T. W., "Packaged Electric-Tuned Panoramic Receiver For 35-200 Mc Range," *Proc. NEC*, 1957.

20. Sharpe, C. B., and D. S. Heim, "A Ferrite Boundary-Value Problem in a Rectangular Waveguide," *IRE Trans. MTT-6*, No. 1, pp. 42-26, January 1958.

21. Sharpe, C. B., "A Graphical Method for Measuring Dielectric Constants at Microwave Frequencies," *IRE Trans. MTT-8*, No. 2, pp. 155-159, March 1960.

22. Sharpe, C. B., and C. G. Brockus, "A Method for Measuring the Dielectric Constant of Ferroelectric Ceramics at S-Band Frequencies," *J. Am. Ceramics Soc.*, Vol. 43, No. 6, June 1960.

23. Sharpe, C. B., and D. S. Heim, "Reflections in a Ferrite Filled Waveguide," *EDG Tech. Report* No. 72, The University of Michigan, Ann Arbor, Mich., May 1957.

24. Diamond, H. and L. W. Orr, "Interim Report on Ferroelectric Materials and Their Applications," *EDG Tech. Report* No. 31, The University of Michigan, Ann Arbor, Mich., July 1954.

25. Orr, L. W., "Ferromagnetic and Ferroelectric Tuning," *EDG Tech. Report* No. 32, The University of Michigan, Ann Arbor, Mich., July 1954.

26. Orr, L. W., "E-T-E Surfaces of Ferroelectric Ceramics," *EDG Tech. Report* No. 53, The University of Michigan, Ann Arbor, Mich., October 1955.

27. "Addition of Fe_2O_3 to $BaTiO_3$—$SrTiO_3$ Ferroelectric," Solid States Devices Laboratory, Department of Electrical Engineering, The University of Michigan, Ann Arbor, Mich., *Technical Report* No. 3, Astia Document, No. AD 132-466. December 1957.

28. "The UHS Spectra of Single Crystals of $BaTiO_3$," *Sylvania Technical Memo* No. EDL-M135, Contract No. DA-36-039 SC-73170, April 1958.

29. "Microwave Phase Shifting with Ferroelectric Elements," *Sylvania Technical Memo* No. EDL-M134, Contract No. DA-36-039 SC-73170, April 1955.

30. Diamond, H., *Polarisation, Microwave Dispersion, and Loss in High-Permittivity Ferroelectrics*, Willow Run Laboratories, The University of Michigan, Ann Arbor, Mich., Report No. 2900-121-T, January 1960.

29

Antennae

J. V. N. GRANGER

29.1 Introduction

Almost all ECM systems require antennas, and it is frequently the antenna requirements which pose the most difficult problems in the design and application of ECM systems. This chapter deals with the nature of the antenna problems encountered in ECM systems and with some of the solutions which have been found for these problems.

As a beginning, we will define certain antenna parameters and briefly describe their interrelation. In the sections which follow, we will discuss the fundamental physical limitations of antenna behavior; describe the kinds of antenna performance which would be optimum for different kinds of ECM systems in various tactical situations; survey some of the different antenna designs which have proved useful in ECM applications; and conclude with a brief description of the kinds of problems which arise in installing antennas.

An antenna is a kind of transducer, which serves to transform electromagnetic energy between those modes which will radiate in space and those modes which occur on transmission lines or in lumped circuits. As such, antennas have both circuit and field properties, *i.e.*, impedance- and radiation-pattern properties. The description and analysis of these properties are considerably simplified by the fact that they are the same whether the antenna is used for receiving or transmitting.*

*This statement is not true in those special cases where the antenna structure includes materials whose properties are nonreciprocal or anistropic, but such situations are rarely encountered in practice.

As with any two-terminal device, the *input impedance* of an antenna is defined as the (complex) ratio of terminal voltage to terminal current when the antenna is connected to a sine-wave generator at a particular frequency. Since most ECM antennas are not connected directly to the associated transmitter or receiver circuits, but rather through a length of transmission line, their impedance properties are frequently expressed in terms of a related parameter, the VSWR or *voltage standing-wave ratio*. The VSWR is defined as the ratio of the maximum and minimum voltages appearing across the line (at points $\lambda/4$ apart) of given characteristic impedance when the antenna terminates the line and the line is energized by a sine-wave generator of specified frequency.

Turning now to the field properties of an antenna, we first define the *radiation pattern*. The radiation pattern of an antenna is a two-dimensional plot of the variation of the antenna response (voltage developed across a fixed load impedance) as the antenna is rotated about a specified axis (fixed both with respect to the antenna and with respect to the incident field) in an incident electromagnetic field of a specified polarization. Since the response of most antennas exhibits a three-dimensional variation, more than one radiation pattern is usually required to completely describe its radiation properties. An *isotropic antenna* is an idealized, lossless (and fictitious) antenna whose radiation pattern is a circle for any orientation of the axis of the pattern. It is useful for reference purposes.

The gain of an antenna is a very useful property in system-performance calculations. It is necessary to distinguish between several meanings of the term. The *power gain* of an antenna is defined as the ratio of the power delivered by the antenna (to a matched load) to the power which would be delivered (to a matched load) by an isotropic antenna located at the same point in the incident field. The *directive gain* of an antenna is defined in the same fashion as the power gain, except that the antenna is assumed to be free of losses. The directive gain can be obtained from the radiation patterns of the antenna, but the power gain requires a knowledge of the antenna efficiency, and is usually obtained by comparative measurements with a standard antenna. In antennas whose radiation patterns are characterized by a single beam (or at least by a beam which is large compared with any side lobes), the directive gain can be calculated approximately from the half-power beamwidths in two orthogonal planes through the beam maximum. Let W_1 be the beamwidth in one plane and W_2 the beamwidth in the other. The beam can then be described as occupying $W_1 \times W_2$ square degrees. The total number of square degrees in a sphere is approximately 41,000. The directive gain with respect to an isotropic source is therefore

$$G = \frac{41,000}{W_1 \times W_2} \tag{29-1}$$

For antennas having more complex radiation patterns the directive gain can be obtained by graphical (or numerical) integration of the radiation patterns (Reference 1). It is useful to refer to the *aperture*, sometimes called the *collecting aperture* or *receiving cross section*, of an antenna. This concept arises from the frequent practice of describing the field incident on an antenna in terms of its power density, in watts per square meter. The aperture is defined in terms of the power gain by the relation

$$A = \frac{\lambda^2 G}{4\pi} \tag{29-2}$$

where, of course, both A and G are regarded as functions of the variables which describe the orientation of the antenna in the incident field. Both A and G in Eq. (29-2) are defined in terms of power delivered to a matched load. It is convenient to define *effective gain* and *effective aperture* (or effective receiving cross section) as the values observed when the load is mismatched by some (known) amount.

It should be noted that the aperture of an antenna is not in general equal to its physical cross section. Particularly when describing reflector or horn-type antennas whose cross-sectional dimensions are large compared to the wavelength, it is customary to speak of the *aperture efficiency*, defined as the ratio of the aperture to the cross section. The aperture efficiency usually falls between 50 and 60 percent in practice.

The utility of the antenna parameters just defined can be illustrated by an example. Suppose we wish to calculate the power received on a given antenna from a transmitting antenna at a distance, both assumed to be in free space. The appropriate equation is

$$P_R = P_T \frac{G_T G_R \lambda^2}{16\pi^2 R^2} \tag{29-3}$$

for matched loads. Here P_R and P_T are the received and transmitted powers, respectively; G_R and G_T are the corresponding antenna gains, taking into account the relative orientation of the antennas; and λ and R are in the same units.

In terms of aperture, Eq. (29-3) becomes

$$P_R = P_T \frac{A_T A_R}{\lambda^2 R^2} \tag{29-4}$$

The modification of these formulas which is required when the propagation involved is not "free space" is discussed in Chapter 31.

29.2 Antenna Bandwidth: Definitions and Limitations

Because of the ever-present pressure in ECM systems to increase coverage to higher and to lower frequencies, and to accommodate the widest possible band of frequencies in each system component, one of the most pertinent basic considerations in ECM antenna design is the question of the inherent bandwidth capabilities of antennas of limited size. Every engineer who has ever undertaken the design of an ECM antenna is aware of the practical difficulties involved in achieving antennas combining patterns and impedance which are useful over wide frequency bands with acceptable antenna size. Those who have undergone this experience are uniformly convinced that nature sets fundamental limits on the bandwidth which can be achieved with an antenna structure of a given size, and indeed this is true.

In discussing the bandwidth of an antenna, we must first define what is meant. We do not mean the "tuning range" of the antenna. However useful antennas that require readjustment of their physical parameters when the input frequency is changed may be in practice, such schemes are excluded from consideration here. In the discussion which follows we confine our attention to antenna structures whose physical parameters (including the physical values of the elements of the matching network) are fixed. We define the *bandwidth* of an antenna, then, as the frequency range over which all of the operationally significant performance parameters remain within acceptable limits. We will generally deal with the fractional bandwidth, β, defined by

$$\beta = \frac{(f_2 - f_1)}{\sqrt{f_2 f_1}} \tag{29-5}$$

where f_2 and f_1 are the upper and lower limits, respectively, of the frequency range through out which the performance is acceptable.

The requirement that *all* of the important performance parameters be acceptable is so obviously reasonable that it hardly seems necessary to labor the point. In fact, however, it is not uncommon for antenna inventors, and even antenna users, to restrict their attention to only one parameter, usually the input impedance, when describing the bandwidth of a particular antenna. The consequence of this regrettable single-mindedness can be, and too often is, an ECM system which fails to fulfill its intended purpose.

Defining an acceptable range of input impedance is relatively easy. It is usually done in terms of the VSWR on a transmission line of established characteristic impedance. Traditionally, VSWR's of 2:1 or less (correspond-

ing to a maximum power reflection of 11 percent) are regarded as acceptable for transmitters employing negative grid tubes. Magnetron transmitters, which are subject to frequency pulling when connected to mismatched long lines, usually require lower VSWR's. In receiving applications, VSWR's up to 5:1 are generally acceptable.

Defining acceptable variations of the radiation patterns of ECM antennas is a much more sophisticated matter. Radiation patterns which are optimum for various operational situations are discussed in the Sections 29.3 and 29.4. An optimum pattern represents a prescription of the gain over the angular range of interest. From this standpoint, then, a definition of acceptable pattern performance would take the form of allowable deviations from the optimum gain. In addition to the complexity inherent in the multi-dimensioned nature of such a criterion, a further complication arises because the variation of ECM system effectiveness with antenna gain is not linear, so that a decibel in one sector of the radiation pattern "costs" more (or less) than a decibel in some other sector. In practice, it is customary to set limits on the radiation pattern in terms of acceptable ranges of power gain within a specified angular sector, disregarding the pattern outside this sector.

In addition to input impedance and radiation pattern, important parameters in any antenna application, other performance parameters may require specifications in some applications. The polarization performance can be specified in several ways, the most common being to define an acceptable range for the ellipticity ratio or to define the desired polarization and place a limitation on the fraction of the radiated power which appears in the undesired polarization. Where the radiation phase is important, as in some direction-finding systems, acceptable tolerances on this parameter must be established also.

The bandwidth of any physical antenna, defined as above, is subject to certain intrinsic limitations. No radiating structure, whatever its size or configuration, can provide useful impedance, and polarization performance over an indefinitely large bandwidth. Indeed, as we shall see below, the bandwidth becomes vanishingly small if we specify performance exceeding some "natural" value associated with the size of the radiating structure. Super-performance in gain, uniformity of input impedance, and purity of polarization are mathematical curiosities of absolutely no practical value to the antenna designer.

Any antenna, whatever its configuration, is in effect a mode transducer whose purpose is to transform electromagnetic energy from the TEM mode of the feed line or driving circuit to the radiating TE_{nm} and TM_{nm} spherical waves, or vice versa. Associated with this transformation are a large number of other modes all of which serve to store energy, but none of which produce radiation at a distance or contribute usefully to the associated circuits.
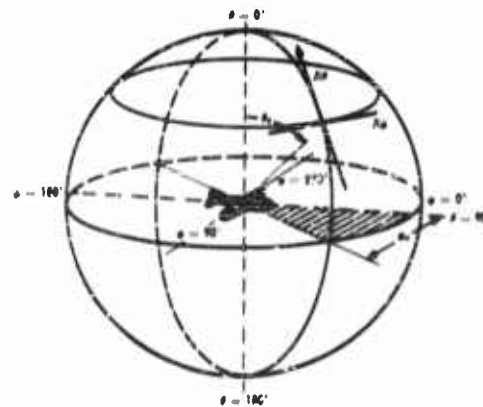
FIGURE 29-1   Coordinate system.
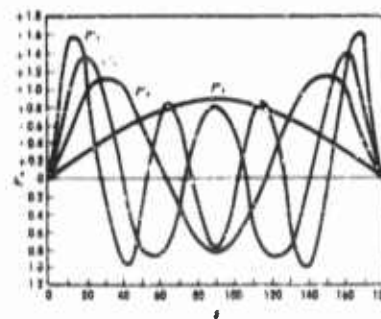
FIGURE 29-2 θ-dependence of elementary radiating modes.

In terms of the coordinate system defined in Figure 29-1, the radiation pattern of every vertically polarized antenna can be regarded as a linear superposition of TM_{nm} modes and that of every horizontally polarized antenna as a superposition of TE_{nm} modes. Both TE and TM modes are present when the radiated field is elliptically polarized. The angular dependence of the elementary spherical waves of both types is described by the series of the associated Legendre polynomials, which behaves somewhat like a Fourier series in the angular interval from 0 to π. The patterns of omnidirectional antennas, for which the radiated field has no ϕ dependence, are described by TM_{n0} and/or TE_{n0} modes. The θ-dependence of the first few of these modes is shown in Figure 29-2. It is clear that, the higher the gain of the antenna, the higher the order of the TE or TM wave involved.

From the impedance standpoint, an antenna can be represented by an equivalent circuit like that shown in Figure 29-3. In this equivalent circuit the coupling networks, $N_1, N_2, \ldots, N_n$ represent the role of the antenna and feed structure in energizing the required radiating modes represented by the impedances $Z_1, Z_2, \ldots, Z_n$. When the radiation pattern is specified, by fixing all of the nonzero I_m's, the antenna design problem is to obtain, by the proper choice of antenna configuration and feeding arrangement, the proper form

of N_m's. In so doing, the designer finds that he has constrained the input impedance. Conversely, he finds that, if the input impedance is initially specified, the radiation pattern will be constrained. It is not possible to design an antenna to meet, simultaneously, independently specified pattern and impedance behavior, even when each is known to be physically realized when associated with other impedance or pattern functions.
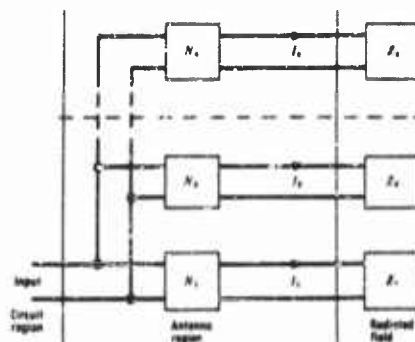
FIGURE 29-3 Equivalent circuit representing antenna.
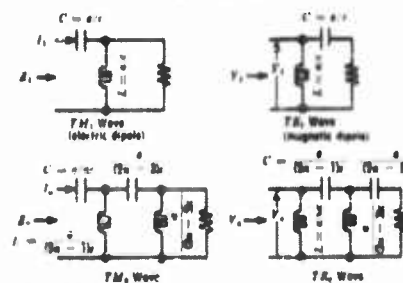
FIGURE 29-4 Equivalent circuits for Z_m's. a = radius of sphere enclosing antenna; c = velocity of light.

The constraints imposed on the input impedance of an antenna with a specified radiation pattern have been analyzed by Chu (Reference 2). He describes the constraints on the input impedance in terms of lower bounds for the Q of the antenna. To obtain the Q of an antenna with a specified pattern behavior, he describes the Z_m's in the equivalent circuit of Figure 29-3 by appropriate equivalent circuits, as shown in Figure 29-4. He assumes that the networks, N_m, are of the minimum reactance type, that is, that they store only the energy required to make the transformed impedance a pure resistance at the operating wavelength. Since no distributed network, such as an antenna, can truly be of the minimum reactance type, Chu's calculated Q's are lower than can be obtained with physically realizable antennas. His results are nonetheless of great significance in antenna design. The modal equivalent circuits of Figure 29-4 are all high-pass filters. This is a very important observation. Let us re-emphasize it. *Every radiating device is a high-pass structure.* In other words, every physical antenna is characterized by a lower frequency limit below which its radiating properties are useless. When the circumference of the sphere enclosing the antenna is less than n wavelengths, the equivalent circuit for the TE_n or TM_n mode behaves essentially as a pure reactance, *i.e.*, the frequency is below cutoff. For smaller antennas at this frequency, or for lower frequencies, the Q rises astronomically. We note that, for the higher-order modes, corresponding to higher

antenna gains, the cutoff frequency is directly proportional to the order of the mode. In other words, the greater the gain required at a given frequency, the larger the antenna structure required. It is ol lous that the input Q of the antenna is higher when the higher-order modes are used. This accounts for the important but sometimes insufficiently recognized fact that high-gain antennas inherently have narrower bandwidths than low-gain antennas.

It is obvious that the higher the Q of an antenna, the more difficult it is to match over a wide band. In fact, if the degree of match required is too stringently defined, a high-Q antenna becomes a spot-frequency device no matter how sophisticated the impedance-matching network. The general problem has been studied by Fano (Reference 3). Similar results in a particularly useful form have been given by Tanner (Reference 4). These are summarized in Figure 29-5. This curve shows the optimum matching of an antenna whose input impedance can be represented over the frequency range of interest by a simple resonant circuit, corresponding to Chu's lowest-order modal equivalent circuit, obtainable with a matching network of n elements. The abscissa is labeled in terms of the product βQ, the fractional bandwidth times the antenna Q. The Q of the antenna is obtained from the definition $1/Q = \triangle f/f_r$, where f_r is the resonant frequency and $\triangle f$ the half-power bandwidth. For a simple series-resonant circuit, the half-power frequencies are those for which the resistance is equal to the reactance. The interpretation

Figure 29-5 Standing-wave ratio vs. βQ. β = fractional bandwidth in terms of geometric mean frequency $(f_2 - f_1)/\sqrt{f_2 f_1}$

in the case of measured antenna impedances can be visualized in terms of Figure 29-6. Here the measured impedance of a particular antenna is shown as the dashed curve on a Smith chart. Rotated through a section of transmission line, the impedance takes the form of the solid curve with circle points. This impedance can be quite accurately represented by the equivalent circuit of Figure 29-7. When the shunting susceptance of the parallel resonant circuit has been subtracted, the impedance follows the curve defined by the square points. Inspection of this curve shows that the half-power frequencies are approximately 216 and 230 Mc, and the resonant frequency is approximately 222 Mc. Thus

$$Q = \frac{222}{230 - 216} = 15 \qquad (29\text{-}6)$$

Fano has shown that the optimum matching of an impedance of this type is obtained with a network of the form shown in Figure 29-8, consisting of alternate shunt parallel-resonant and series series-resonant circuits. When the equivalent circuit of the impedance to be matched already includes a shunt-
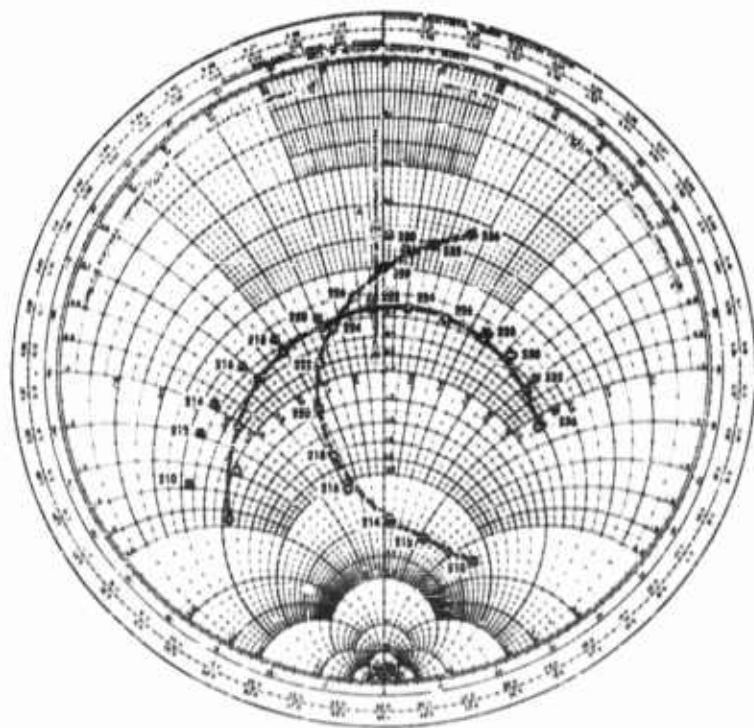
FIGURE 29-6   Measured antenna impedances.

FIGURE 29-7   Equivalent circuit, antenna impedance of Figure 29-6.
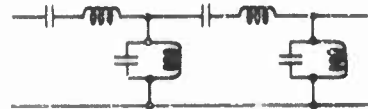
FIGURE 29-8   Matching network configuration.

ing circuit, this susceptance can be considered as being lumped with the first element of the matching network. An optimum match can be obtained if the built-in susceptance is smaller than that required for matching. If this is not the case, the bandwidth obtainable for a prescribed VSWR tolerance will be smaller than that given by Figure 29-5. This situation will be recognized by antenna designers as the case where the impedance curve is already "tied too tightly" around the center of the Smith chart. The reader with experience in impedance matching will recognize also another result of Fano's work, that is, that the widest bandwidth at a given VSWR tolerance will be obtained with a matching network which does not yield a perfect match at any point in the band.

The limitations on impedance bandwidth given in Figure 29-5 are fundamental. Other matching procedures can be used, of course, but the results obtained will, at best, equal those given in the figure.

After the above discussion we can list some of the features which are common to all antennas capable of wide bandwidth. Their patterns are simple and the gain is relatively low. Their structure is simple and "clean," with a minimum of sharp corners and other regions for excess stored energy. At least one dimension is of the order of $\lambda/2$ at the lowest frequency in the band. Truly broadband antennas do not require external matching elements, and are not perfectly matched at any point in the band.

Section V describes briefly a variety of antennas which have such characteristics.

29.3 Optimum Antenna Patterns—Jamming*

In Chapter 13 the problem of determining the amount of jamming power required to disable a radar is explored. In this section we will explore the implications of those results in establishing the radiation pattern requirements on a jamming antenna system. We will deal here with the ratio J/S, where J is the jamming power density and S the signal-power density, measured at the radar antenna. Expressed in decibels, the general equation for J/S, is

$$\frac{J}{S} \text{ db} = 20 \log_{10} R + G_j \text{ db} + P_j \text{ db} - 10 \log_{10} \sigma - \frac{1}{4\pi}\left(\frac{P_t G_r}{B_r}\right)\text{db (29-7)}$$

where R is the slant range between the jammer and the radar (meters)

G_j is the antenna gain of the jammer

P_j is the power density of the jammer (watts/megacycle)

*The material presented in this section is largely adapted from the excellent study of Reference 5.

σ is the radar cross section of the protected vehicle (square meters)

P_r, G_r, and B_r are the power output, antenna gain, and effective bandwidth of the radar, respectively

This equation assumes that the jammer is carried on the vehicle to be protected and that the output spectrum of the jammer is in the form of white noise. The ratio P_rG_r/B_r expresses the role of the radar system parameters in determining J/S. For typical modern tracking radars, the ratio has values lying between 100 and 120 dbw/Mc. For surveillance radars, the ratio is dependent on the elevation angle, typically falling about 6 db from the horizon to 20 degrees above the horizon, with horizon values for typical modern radars of 90 to 113 dbw/Mc.

In order to explore the effect of the jammer antenna pattern on J/S, we must first obtain a value of σ, the cross section of the vehicle to be protected. As is well known, static measurements of radar cross section show extreme fluctuations in magnitude which appear in a moving target as scintillation. The median value of the cross section in 10-degree sectors is convenient for our purposes here. For a B-47 aircraft, for example, a good approximation for the median value of σ is independent of the elevation angle, and has quadrant symmetry in azimuth. It is described in the equation

$$\sigma = 10 \text{ m}^2 \qquad (0 \le \phi \le 60°)$$
$$= 10^{-3}e^{8.84} \text{ m}^2 \quad (60 \le \phi \le 90°)$$

The radar cross section of ship targets is complicated by the fact that such targets often intercept more than one lobe of the radar beam, and may have dimensions (in range) exceeding one pulse length. Average values have been obtained by comparison measurements at the Naval Research Laboratory. According to these data (Reference 6), the following are typical at X-band:

Vessel	db above 1 m²
Battleship	66
Aircraft carrier	61
Heavy cruiser	56
Destroyer	48
Destroyer escort	46

Inserting the appropriate values for σ, for P_j, and for the ratio P_rG_r/B_r, we can calculate the spatial distribution of J/S from G_j. Conversely, if we can determine, by the methods outlined in Chapter 13, the required values of J/S at all possible radar locations, we can determine the optimum spatial variation of G_j, i.e., the optimum jammer antenna pattern. We will illustrate the first procedure with an example. Consider an airborne jammer at an alti-
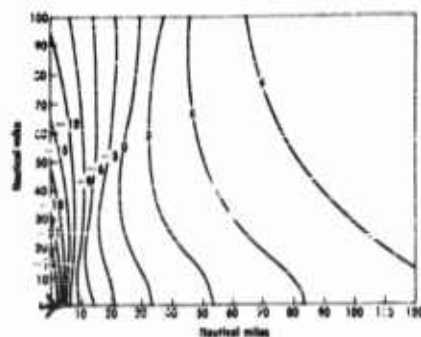
FIGURE 29-9 Calculated contours of constant J/S. Contours marked with J/S in decibels.
B-47 cross section approximation A. Jamming power 100 w/Mc against a Nike-Zeus-type tracking radar. Aircraft altitude 75,000 feet. Linear polarization.
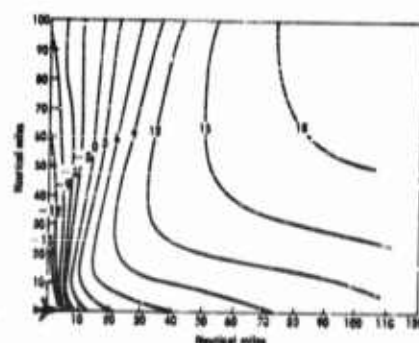
FIGURE 29-10 Contours for jamming antenna system of four sectoral horns, radar known to be in quadrant shown. Contours marked with J/S in decibels. Aircraft altitude 75,000 feet. B-47 cross-section approximation A. Jamming power 100 w/Mc against Nike-Zeus-type tracking radar. Jamming antenna beamwidth $= 65°$. Gain $= 10$ db. Directed at $\phi_d = 45°$; $\theta_d = 120°$. Linear polarization.
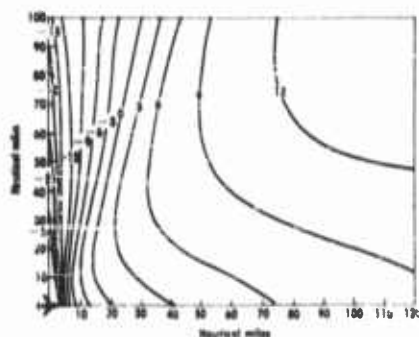
FIGURE 29-11 Contours for jamming antenna system of four sectoral horns, all antennas connected in parallel. Contours marked with J/S in decibels. Aircraft altitude 75,000 feet. B-47 cross-section approximately A. Jamming power 100 w/Mc against a Nike-Zeus-type tracking radar. Jamming antenna beamwidth $= 65°$. Gain $= 10$ db. Directed at $\phi_d = 45°$, $135°$, $225°$, and $315°$. $\theta_d = 120°$. Linear polarization.
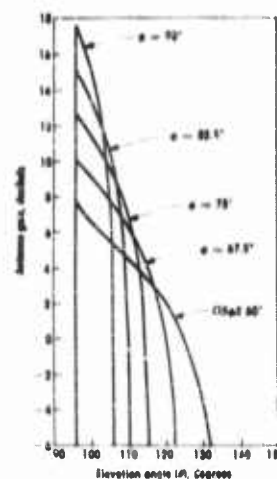
FIGURE 29-12 Optimum jammer antenna patterns. Antenna gain in decibels above isotropic gain. Aircraft altitude 75,000 feet. Jamming to be limited to radius of 133 miles. Pattern symmetry in all four quadrants assumed. Linear polarization.
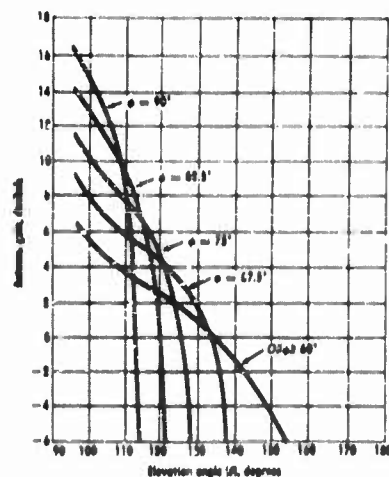
29-12

FIGURE 29-13 Optimum jammer antenna patterns, jammer power density three times higher than for Figure 29-12. Aircraft altitude 75,000 feet. Jamming to be limited to radius of 123 miles. Pattern symmetrical in all four quadrants. Linear polarisation.
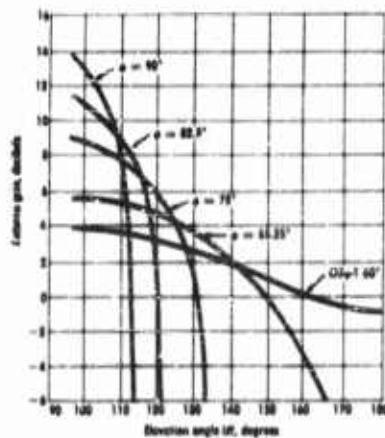
FIGURE 29-14 Optimum jammer antenna patterns, kill probability inversely proportional to range. Antenna gain in decibels above isotropic gain. Aircraft altitude 75,000 feet. Jamming to be limited to radius of 123 miles. Pattern symmetry in all four quadrants assumed. Linear polarisation.

tude of 75,000 feet with a power output of 100 w/Mc flown against a tracking radar with a value for the ratio P_rG_r/B_r of 118 dbw/Mc. Figure 29-9 shows the calculated contours of constant J/S on the surface of a plane earth in the quadrant forward and to the left of the aircraft, for a linearly polarized isotropic jamming antenna. Figures 29-10 and 29-11 show similar sets of contours for the case where the jamming antenna system consists of four sectoral horns, each with a beamwidth of 65 degrees and a gain of 10 db, directed 30 degrees below the horizon and at azimuth angles of 45, 135, 225, and 315 degrees. In Figure 29-10 it is assumed that the general location of the radar is known to be in the quadrant shown, and all of the power has been switched to the appropriate antenna. In Figure 29-11 all of the antennas are connected in parallel, but the outputs are not correlated, in order to eliminate the interference lobing between adjacent antennas. (This can be achieved by wide spacing of the antennas or by inserting in the feed of one of two adjacent antennas a time delay which is long compared with $1/B_r$.) Contour plots of this type can be utilized to obtain quantitative comparisons of the jamming effectiveness of different antennas, if the characteristics of the radar are known. We can assume a value for J/S at which jamming is effective, and compute the fraction of the ground area surrounding the aircraft and within the radar horizon over which the threshold value of J/S is equaled or exceeded. For several reasons, some of which are discussed in

Chapter 13, such a comparison is of limited interest. A more meaningful comparison can be obtained by weighting the coverage areas obtained at each J/S in accordance with an appropriate jamming effectiveness function (which will always be monotonic if the radar involved is at all sophisticated) and summing as before.

The procedure for obtaining the optimum jamming antenna patterns consist of expressing in appropriate analytic form the effectiveness of the jamming antenna as described above, and of then obtaining, by an application of the calculus of variations, the functional form of G_j which maximizes the jamming effectiveness expression. For the details of the r⁻ be tricky mathematical procedures involved, the reader is referred to Section IVe of Reference 6. Two examples will be shown. Figure 29-12 shows the optimum jammer antenna patterns for the operational circumstances described previo. In this calculation the jamming is limited arbitrarily to a radius of 123 miles in order to avoid the singularities introduced in the integrals by the use of a plane earth model. The effect of the broadside peaks in σ is obvious. It is apparent also that radiation in the sector immediately below the aircraft is not significant to jamming effectiveness. Figure 29-13 shows the optimum patterns for the same situation except that the jammer power density is here assumed to be three times higher (or the value of the ratio P_rG_r/B_r is 4.8 db lower) than in Figure 29-12. The most significant change in the patterns is the reducton of the size of the "no jamming" null below the aircraft.

The optimum pattern analysis described has at least two weaknesses. The first is that the derivation assumes a knowledge of the radar performance, and in particular, the relationship between J/S and jamming effectiveness. In practice, it is virtually impossible to determine these characteristics about our own radars, let alone those of a potential enemy. Some relief from this dilemma is afforded by the fact that analysis shows that the optimum patterns calculated from an assumed set of radar parameters are very nearly optimum for variations up to ± 6 db in radar performance. The second weakness of the optimization procedure is that it does not take into account the fact that in many cases jamming the radar is only a means to an end; it is degrading the kill probability of the weapon system which we are really after.

The relationship between radar jamming effectiveness and weapon-system kill probability involves an array of intricate factors, most of which are not determinable. It is worth while, however, to examine the effect of introducing various reasonable assumptions. For example, the kill probability of certain ground-to-air missile systems is a function of distance from the launch point to the target. In Figure 29-14 is shown the optimum jammer-antenna patterns for the same situation as in Figure 29-13, except that we have assumed

here that the kill probability is inversely proportional to range, and that the radar and the missile launcher are located at the same site. Note that the new pattern distributes the jamming power so that less energy is directed toward the horizon, and more energy is directed downward, as would be expected.

The threat from surface-to-surface, ground-to-air, or air-to-air missiles will have definite range and velocity limitations which will be primary factors in determining the   boundary of the space in which jamming is significant. If the target were stationary in space, the range limitation of the missile would eliminate the threat from missile launchers outside a circle having a radius about equal to the maximum range of the missile, centered on the target. In the case of aircraft targets, the center of the circle of vulnerability lies forward of the aircraft a distance which depends on the relative velocity of the aircraft and the missile. When the two velocities are equal, the circle of vulnerability will pass through a point just forward of the aircraft, since a missile launched just after the aircraft passes will never overtake the aircraft. Analytically, the center of the circle of vulnerability is centered at a point forward of the aircraft a distance d, where

$$d \sim R(v_a/v_m)$$

and a radius R, where R is the maximum range of the missile, v_a and v_m are the average velocities of the aircraft and missile, respectively,* and the approximation sign is necessary because of altitude and maneuverability considerations.

In evaluating jamming antenna patterns, the existence of a circle of vulnerability is an important factor, since jamming power radiated outside it is ineffective. Sometimes other factors associated with the weapons system modify the optimum spatial distribution of jamming energy. For example, some radars cannot be slewed fast enough to track targets passing nearly overhead. Operational doctrine may limit the sector of vulnerability, as, for example, when tail-chase missile launchings are prohibited because of the danger from the tail guns of the target aircraft, or when targets are passed through a radar net in such a way that the effectiveness of antijam techniques designed to operate against chaff are optimized.

Time delays in the ground defense system can have a critical effect on the optimum distribution of jamming energy. For example, after surveillance and acquisition radars have clearly detected and identified a target, the

*For air-to-air missiles, the missile velocity includes a vehicle velocity and a differential missile velocity at launch, so that the average missile velocity for small flight distances is high.

target-tracking radar must be slewed to the correct azimuth and elevation, and be permitted to track continuously for some minimum time to permit computer "settling"; the total delay involved may be of the order of 15 seconds, during which a Mach 3 aircraft has traveled 8 miles. The time of flight of the attacking weapon may represent a substantial time delay. If the target must be continuously tracked by the ground radar for the full time-of-flight, the effective J/S may be taken as the highest J/S experienced by the radar during that interval. Figure 29-15 shows a typical set of J/S contours
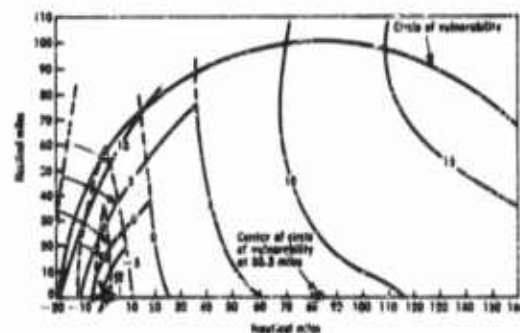
FIGURE 29-15 Typical set of J/S contours adjusted to account for effect of time delay. J/S contours marked in decibels. Shift of J/S contour lines to account for minimum tracking time equal to missile flight time from launcher to intercept point. v_a = Mach 2.5; v_m = Mach 3.0; R_m = 100 miles.

adjusted to account for the effect of time delay. The shifts in the J/S contours broadside and to the rear of the aircraft are marked, and would substantially affect the optimum jamming antenna pattern.

29.4 Optimum Antenna Patterns—Intercept

In this section we will employ an analysis similar in structure to that used in the previous section to obtain the optimum radiation pattern for an antenna to be used with an intercept receiver. The appropriate measure of the performance of an intercept receiver is its output signal-to-noise ratio for a specified radar and a specified tactical geometry. Expressed in decibels, the general equation for S/N is

$$S/N = + 20 \log_{10} \lambda - 20 \log_{10} R + P_n \text{ dbw} + G_n \text{ db}$$
$$+ G_i - NF - 10 \log_{10} B_i + 122 \text{ db} \qquad (29\text{-}8)$$

where λ is the wavelength (meters)

R is the slant range between the intercept receiver and the radar (meters)

P_R is the peak power of the radar

G_R is the antenna gain of the radar

G_i, NF, and B_i are the antenna gain, noise figure, and effective bandwidth of the intercept receiver, respectively

This equation assumes line-of-sight propagation. The value of B_i is calculated from the expression

$$B_i = \sqrt{2B_{rf}B_v - B_v{}^2}$$

where B_{rf} and B_v are the rf and video bandwidth, respectively, and all the terms are in megacycles.

It is worth while to calculate the spatial distribution of S/N for typical cases. We will assume that a crystal video intercept receiver with traveling-wave-tube preamplification is carried in an aircraft at an altitude of 75,000 feet, and calculate the variation of S/N with ground distance for an isotropic intercept antenna. Figure 29-16 shows two cases, as follows:

FIGURE 29-16 S/N vs. ground distance for side-lobe reception. Curves labeled for parameter given in text.

CASE I: L-Band (1000 Mc)

$P_R = 250$ kw

$G_R = $ typical of a surveillance radar, varying from 42 to 21 db with elevation angle

$NF = 25$ db

$B_i = 31.6$ Mc, corresponding to 500-Mc rf bandwidth and 1-Mc video bandwidth

CASE II: X-Band (10,000 Mc)

$P_R = 250$ kw

$G_R = 43$ db, corresponding to a 6-ft antenna tracking the aircraft

$NF = 30$ db

$B_i = 150$ Mc, corresponding to 3000-Mc rf bandwidth and 4-Mc video bandwidth

The signal-to-noise ratios plotted in the figure are for side-lobe reception,

assumed to be 30 db below the main-lobe levels. The effect of the elevation pattern of the surveillance radar on the S/N distribution is obvious.

From Figure 29-16, it is clear that antenna gain is not a limitation on the performance of the intercept system against surveillance radars at the low microwave frequencies. On the other hand, the signal-to-noise performance of the assumed intercept system against X-band radars is inadequate at relatively short distances when a low-gain receiving antenna is used. If an intercept antenna with a gain of 10 db were used in the same circumstances, an S/N of 10 db could be obtained at a range of 80 nautical miles, a nearly ten fold increase in the ground area covered by a ferret mission.

If the frequency is increased still further, the necessity for a relatively high-gain intercept antenna becomes even more apparent. Consider a typical situation at 30,000 Mc.

$$P_H = 25 \text{ kw}$$
$$G_H = 45 \text{ db}$$
$$NF = 20 \text{ db (good traveling-wave-tube performance)}$$
$$B_{eff} = 350 \text{ Mc}$$

With an isotropic intercept antenna, the maximum range on -30-db side lobes for $S/N = 10$ db would be about 16 nautical miles. If the intercept antenna had an aperture 2 inches square, yielding a gain of 21 db, the maximum range would be increased to 87 nautical miles.

The extent to which increased intercept antenna gain can be utilized to increase the coverage area of a ferret mission is limited by other considerations. From the standpoint of technical intelligence it is desirable that a target radar remain within the beam of the intercept antenna for several radar scan periods. With high-speed aircraft, high-gain intercept antennas may not permit this.

In some ferret installations, the intercept antenna beam is pointed 45 degrees off the nose, rather than broadside, in order to increase the length of time that a target radar at a particular distance from the track of the aircraft remains within the intercept antenna beam. This procedure has the effect of increasing the minimum range to the target radar, and it is necessary to increase the antenna gain (by 3 db) in order to maintain the same S/N. The increased gain results in a narrowed beamwidth, so that no improvement in "intercept time" is achieved. If the lower S/N can be tolerated it is, of course, possible to increase the intercept time by using a lower-gain antenna pointing broadside.

A truly "optimum" intercept antenna pattern would employ beam shaping in the vertical plane to obtain a uniform S/N performance independent of range. Only if the receiver sensitivity is very low, or the signals of interest

are very weak, does this complication appear worth while, however. A consideration of far greater practical importance is finding intercept antenna locations which yield "clean" patterns, free from deep lobing due to reflections from surrounding structures. Such lobes can seriously deteriorate the performance of the intercept system.

29.5 Typical ECM Antenna Designs

A tremendous array of different antenna types have been developed for ECM applications. A complete survey would be beyond t   scope of this chapter. Rather, we will confine ourselves here to a survey of a number of antenna types which have found important applications to ECM. No attempt will be made to present design data. Our attention will be confined to those principal characteristics of greatest interest to the potential user—VSWR and pattern bandwidth, gain, and over-all size. The categories used are for convenience only, and are not presumed to be entirely unambiguous and nonoverlapping.

Electric Dipoles. Electric dipoles and monopoles are perhaps the most widely used of all antennas. Many different designs are available, but all are characterized by a radiation pattern which is rotationally symmetric about the axis of the antenna, and polarized in the plane containing the axis of the antenna. When the electrical length of the dipole is near or below $\lambda/2$ ($\lambda/4$ for a monopole), the gain is nearly constant at 2 db relative to an isotropic source, and the receiving cross section is approximately $\lambda^2/8$. As the electrical length is increased beyond $\lambda/2$, the gain is increased slowly and the pattern breaks into several lobes.

The VSWR bandwidth of an electric dipole tends to increase as the length-to-thickness ratio decreases. This effect is utilized in the conical monopole and biconical dipole, Figure 29-17a and b. The 2:1 VSWR bandwidth can be as large as 10:1 for a 30-degree half-angle cone, with the low-frequency cutoff at a radius of about $\lambda/6$. The radiation patterns tend to break up when the cone radius exceeds 3λ, however. A modification proposed by J. F. Byrne, Figure 29-17c, uses conical power-dividing septums to suppress the higher-order TM modes and yields a pattern bandwidth comparable with the VSWR bandwidth. The discone antenna shown in Figure 17d has bandwidth properties similar to the simple cone, but does not require a ground plane. Figure 29-17e shows a dipole version of the discone, incorporating a balun, which has an equally good bandwidth and is particularly useful as a feed for a reflector-type antenna.

A simple antenna possessing some of the attractive bandwidth features of the cone but more practical at the longer wavelengths is the multiwire fan. Fans with a half-angle of 30 degrees constructed of four or five co-planar
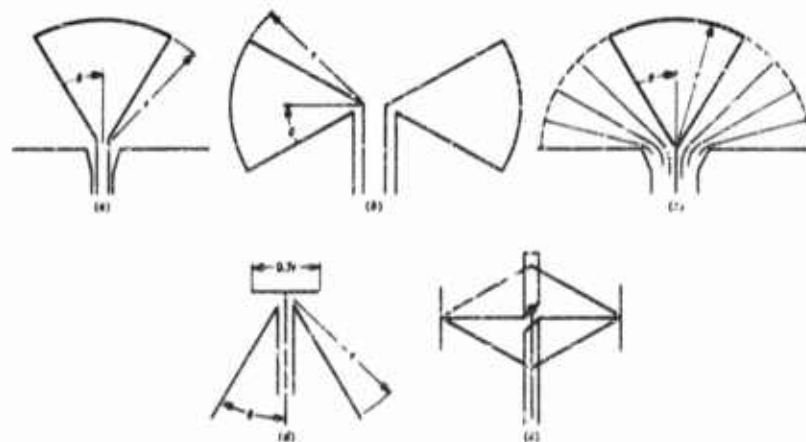
FIGURE 29-17   Dipoles. (a) Conical monopole; (b) biconical dipole; (c) Byrne modification; (d) discone antenna; (e) dipole version of discone
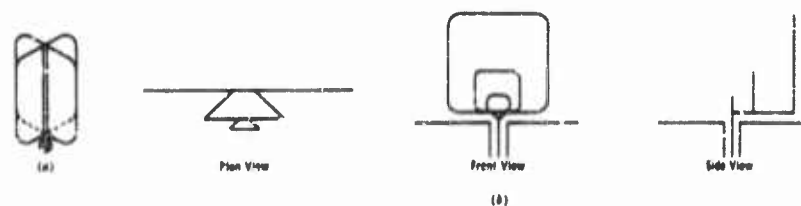
FIGURE 29-18   Planer monopoles.

wires are quite broadband. The VSWR uniformity can be increased somewhat by tapering the length of the wires across the fan.

Planar monopoles* possessing 5:1 VSWR bandwidths of 10:1 are sketched in Figure 29-18. The crossed-plate configuration of Figure 29-18a yields quite uniform H-plane patterns, while the three-plate periodic structure of Figure 29-18b has H-plane patterns which tend to become unidirectional at the higher frequencies due to the reflecting actions of the larger plates. Low-frequency cutoff occurs when the maximum plate dimension is approximately $\lambda/8$.

In the sleeve dipole, one or two auxiliary conductors are arranged so as to transform the feed point to a region of the current distribution which provides broadband impedance compensation. Several variations are sketched in longitudinal section in Figure 29-19. The 2:1 VSWR bandwidth of these

*Developed at Sylvania Electronic Defense Laboratory.
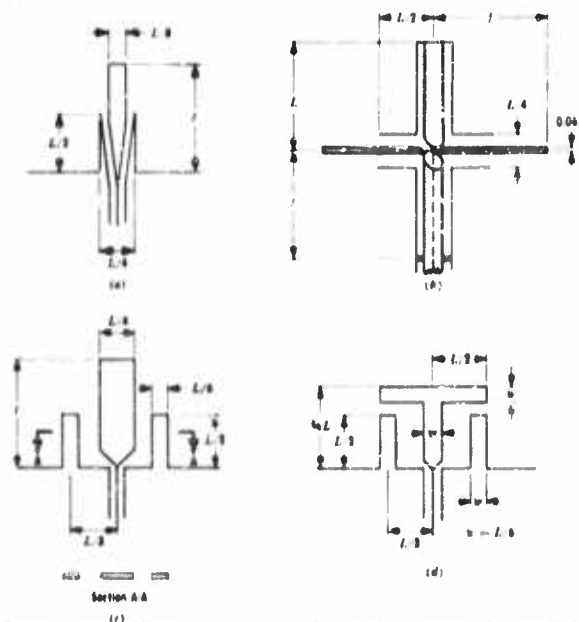
Figure. 29-19   Sleeve dipole configurations, longitudinal sections.

designs can be as great as 5:1 with the low-frequency cutoff at approximately
$L = \lambda/6$. The patterns tend to split up when L exceeds $\lambda/2$, but may be use-
ful to perhaps twice that frequency. The partial sleeve configurations of
Figure 19c and d are particularly adapted to strip conductors as shown. In
this form they are sometimes embedded in reinforced plastic panels to pro-
vide a compact and rugged antenna "package."

The annular slot antenna is a flush-mounted electric monopole which is
useful in certain ECM applications, particularly on aircraft. The pattern
and VSWR can be controlled quite closely over bandwidths of about 2:1, if
the depth of the backing cavity is sufficient. The low-frequency cutoff occurs
when the circumference of the slot is of the order of 1.4λ.

A final electric dipole type worthy of mention is represented by the isolated
wing-cap and tail-cap antennas used in certain aircraft applications at HF
and VHF. The general configuration is sketched in Figure 29-20. A properly
proportioned isolated section can yield reasonably good VSWR bandwidths at
frequencies as low as 30 Mc on an aircraft the size of a B-47. The patterns
are characterized by large numbers of lobes and substantial cross polarization.
Since "clean" patterns are virtually impossible to obtain on aircraft in this
frequency range, the usual practice is to rationalize the utility of the pat-
terns obtained.
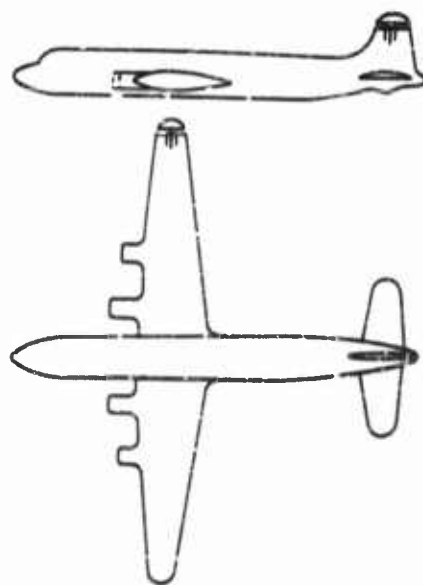
Figure 29-30 Isolated wing-cup and tail-cap antennas.

Magnetic Dipoles. The classic form of the magnetic dipole is the constant current loop, which is almost never used for ECM. A magnetic dipole which is used is the linear slot. Slot antennas which radiate on both sides of the metallic surface in which they are cut have impedance and pattern characteristics which are complementary to those of electric dipoles, so that their bandwidth properties and cutoff dimensions are the same. Where a backing cavity is required, which is usually the case, the bandwidth is reduced.

Three different slot antenna arrangements are sketched in Figure 29-21. The T-fed slot of Figure 29-21a has a 2:1 VSWR bandwidth of about 2:1, with a low-frequency cutoff at $L = 0.6\lambda$. The E-plane half-power beamwidth varies from about 80 degrees at the low end to about 45 degrees at the high end of the band. The arrangements of Figure 29-21b and c* which radiate on both sides of the slot plane have VSWR bandwidths of approximately 3.5:1. The conical probe of Figure 29-21c is particularly suited for arrangements where the slot is cut in a septum projecting from the ground plane as shown and provides a smoother transition to the feed cable than that

*Developed originally at the Stanford Research Institute.
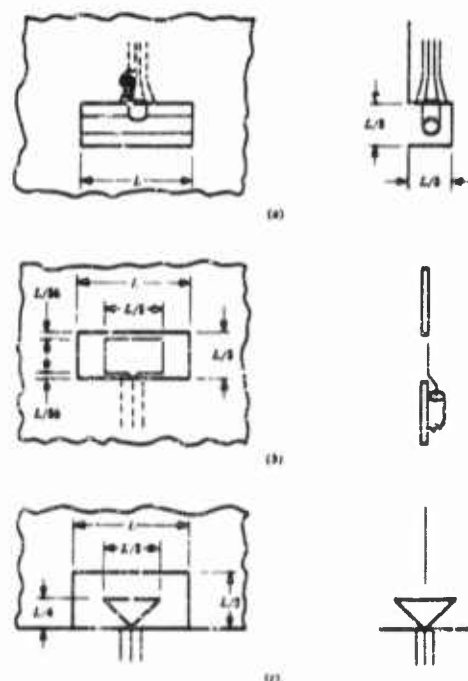
FIGURE 29-21 Slot antennas, front views and cross sections.

of Figure 29-21b at microwave frequencies. Low-frequency cutoff occurs at $L = \lambda/3$ for these arrangements.

Figure 29-22 shows the construction of a slotted cylinder antenna* which provides an omni-azimuthal horizontally polarized pattern and a 5:1 VSWR bandwidth of about 1.7:1. Its low-frequency cutoff occurs when the length of the cylinder is approximately 0.4λ. Figure 29-23 shows the construction of a waveguide-fed slotted cylinder which provides horizontally polarized omni-azimuthal radiation from 4.4 to 9.1 kMc.

Helix and Spiral Antennas. The helix and the spiral are examples of antennas which produce radiation which is inherently circularly polarized. The helix, Figure 29-24a, can radiate in two different modes, the axial mode and the normal mode. In the axial mode, which occurs when the circumference is of the order of the wavelength, the field is a maximum in the direction of the helix axis. In the normal mode, which occurs when the helix dimensions are

*Developed at Airborne Instruments Laboratory.
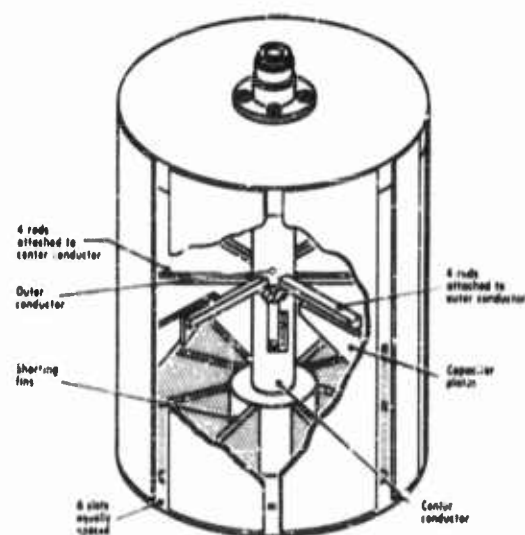
FIGURE 29-22   Horizontally polarized omnidirectional antenna.
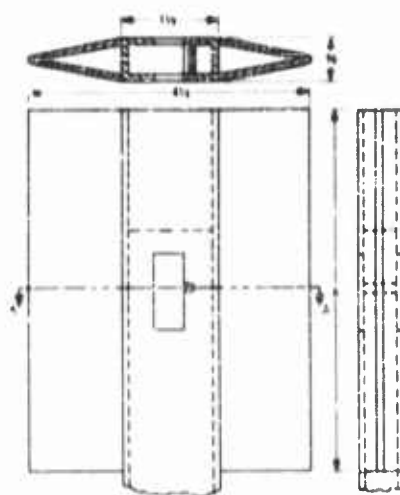
FIGURE 29-23   Blade antenna with slots in broad faces. Dimensions in inches.
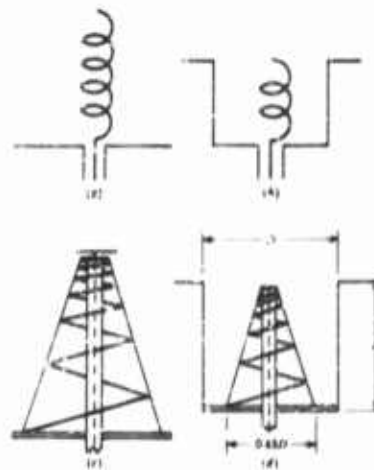
FIGURE 29-24   Helical antennas. (a) Helix; (b) circularly polarized cavity antenna; (c) helix with increased bandwidth; (d) cavity-mounted conical helix

small compared with the wavelength, the field is a maximum in the plane normal to the helix axis. Optimum axial mode performance occurs when the pitch angle is about 14 degrees, the circumference is between $3\lambda/4$ and $4\lambda/3$, and the number of turns exceeds 3. Designs of this kind provide excellent VSWR performance, constant beamwidth and circular polarization over bandwidths of up to 1.8:1, with low-frequency cutoff occurring when the circumference is about $0.7/\lambda$. The gain is a function of the number of turns, and is about 10 db for a 6-turn helix. Short (2- or 3-turn) helixes are often enclosed in circular cavities to provide a flush-mounted circularly polarized antenna, as sketched in Figure 29-24b. In order to maintain the low-frequency cutoff of the helix, the cavity diameter must be approximately 2.4 times that of the helix.

The bandwidth of the helix can be substantially increased by winding the helix on the surface of a cone, particularly if the spacing between the turns is increased logarithmically with the turn diameter. The antenna of this type sketched in Figure 29-24c* has a 3:1 VSWR bandwidth of better than 10:1 with the low-frequency cutoff occurring when the maximum turn diameter is about 0.5λ. The beamwidth varies between 70 and 90 degrees. The cavity-mounted conical helix of Figure 29-24d†, which employs a uniform turn spacing, provides a 2:1 VSWR bandwidth of better than 3:1, with the low-frequency cutoff at about $D = \lambda/2$.

The spiral, which was originally conceived by E. M. Turner and developed extensively at The Ohio State University Research Foundation, the University of Illinois, and Massachusetts Institute of Technology, is a very broadband antenna which is inherently circularly polarized when permitted to radiate on both sides of the surface in which it is mounted, particularly when it is made in the form of a logarithmic spiral (Figure 29-25). This antenna can yield a VSWR and pattern bandwidth as great as 20:1. Low-

FIGURE 29-25 Logarithmic spiral antenna.

frequency cutoff occurs at approximately $D = \lambda/2$. When a backing cavity is used, the bandwidth, particularly with respect to the uniformity of axial ratio, is deteriorated to the order of 2:1.

*Designed at American Electronic Laboratories Inc.
†Developed at Boeing Airplane Company.

Logarithmically Periodic Antennas. The logarithmically tapered conical helix and the logarithmic spiral antennas mentioned in the previous section are examples of logarithmically periodic antennas which are related to the class of frequency-independent structures discovered by V. H. Rumsey. In these structures the geometry is arranged so that the patter and impedance repeat periodically with the logarithm of the frequency. For planar structures this is accomplished by defining their shape such that θ is a periodic function of ln r, where r and θ are the polar coordinates in the plane of the structure. Then if ln τ is the period of ln r, the operation of an antenna of infinite extent would be the same at all frequencies related by integral powers of τ. For the basic structure of Figure 29-26a

$$\tau = \frac{R_{n+1}}{R_n} \qquad (29\text{-}9)$$

If the shape of the structure and the factor τ can be made such that the variation of the pattern and impedance over the period ln τ is small, then this will hold true for all periods, the result being a very broadband antenna. For finite structures, it has been found that the "end effects" produced by the abrupt termination of the periodic structure is small, so that wide bandwidths can be achieved in practice, so long as operation is confined to frequencies above the low-frequency cutoff. For the antenna of Figure 29-26a, the low- and high-frequency limits occur when the longest and shortest "teeth," respectively, are approximately $\lambda/4$.

The antenna of Figure 29-26a* radiates a bidirectional horizontally polarized pattern with approximately equal and constant principal plane beamwidths. The VSWR on a 170-ohm balanced line connected across the vertex is close to unity over the band. The nonplanar structure of Figure 29-26b† behaves similarly with frequency, but yields an endfire pattern, with its maximum along the Y axis, and with a front-to-back ratio on the order of 10 db. The nonplanar wire structure of Figure 29-26c has similar properties.

The structures of Figure 29-27a and b yield circularly polarized endfire patterns with the bandwidth properties previously described.‡ These structures take advantage of a property peculiar to the logarithmically periodic element, that is, that the phase center moves linearly back from the vertex by a distance of one period when the frequency is lowered by an amount corresponding to one period. This permits the 90-degree phasing of two orthogonal linear radiating elements required to achieve circular polariza-

*Conceived by R. H. DuHamel at the University of Illinois.

†Conceived by D. E. Isbell at the University of Illinois.

‡The structure of Figure 29-27a was investigated by Isbell at the University of Illinois, and that of Figure 29-27b by J. W. Schomer at Granger Associates.
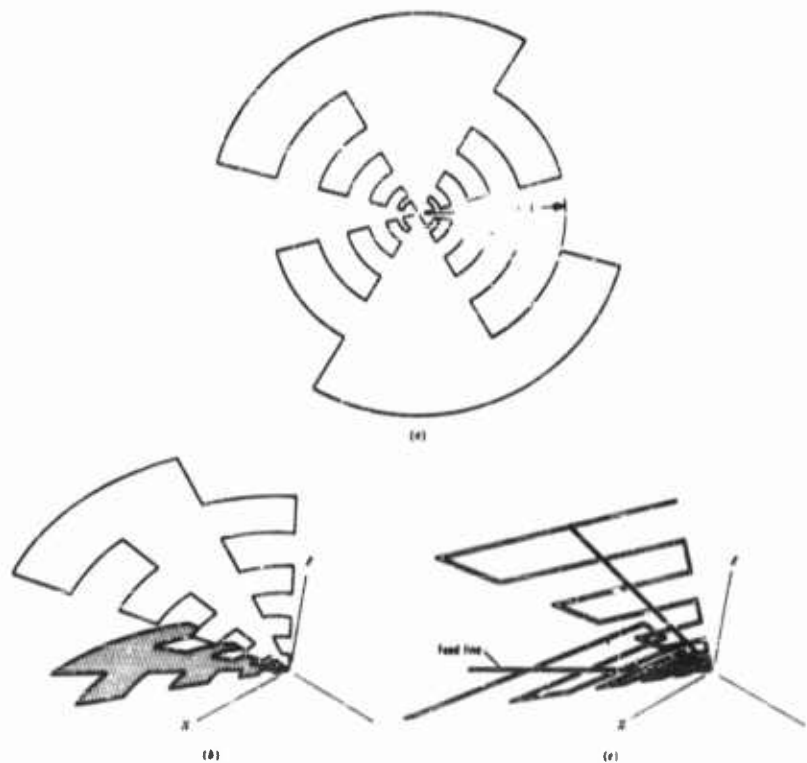
FIGURE 29-26 Logarithmically periodic antennas. (a) Antenna of structure of Eq. (29-9); (b) nonplanar structure; (c) nonplanar wire structure.

tion by constructing the two elements (I and II in Figure 29-27) such that the "space phase" of the structure of one element leads or lags that of the other by one-fourth of the geometric period. A similar procedure is used in the crossed plane structure of Figure 29-28 to obtain an omnidirectional horizontally polarized structure capable of very wide pattern and VSWR bandwidths.

Before leaving the subject of logarithmically periodic antennas, it is worth while to note that high gain endfire patterns exhibiting the frequency-independent properties of those of the antennas mentioned above can be obtained by arraying logarithmically periodic endfire elements like those of Figures 29-26b and 29-27a and b. The arraying required to maintain the frequency-independent property is simply to arrange the elements so that they posses
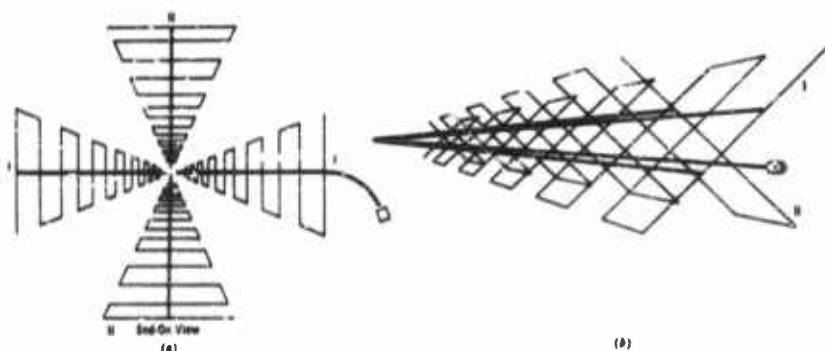
FIGURE 29-27   Structures yielding circularly polarized endfire patterns.

FIGURE 29-29   Transposed dipole logarithmically periodic antenna.

a common vertex.* Mutual imped-
ances are a problem in achieving good
array performance. In this regard, the
transposed dipole logarithmically peri-
odic element† of Figures 29-29 is of
particular interest. This structure is
unique in that while it is approxi-
mately planar it is also endfire. The
absence of radial structures at the
edges of the element cuts down the
mutuals when the element is used in
an array.

FIGURE 29-28   Crossed-plane structure for
omnidirectional radiation.

Horn Antennas. Horn antennas are widely used in ECM applications,
particularly at the highest frequencies, because of their relatively good per-
formance, structural simplicity, and the relative ease with which their proper-

*Such arrays have been extensively investigated by R. H. DuHamel at the Collins Radio
Company.

†Conceived by Isbell.
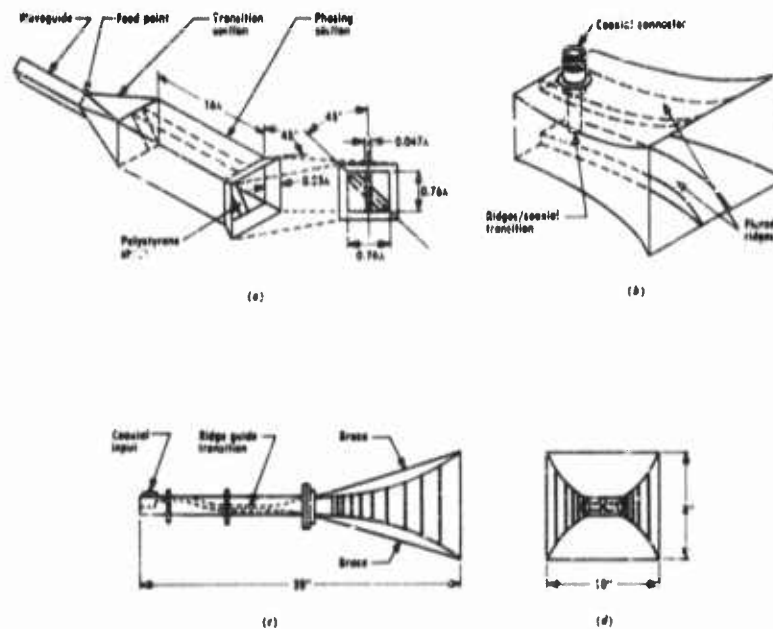
FIGURE 29-30 Horn antennas. (a) Pyramidal horn; (b) linearly polarized horn; (c) and (d) design for approximately constant and equal E- and H-plane beamwidths.
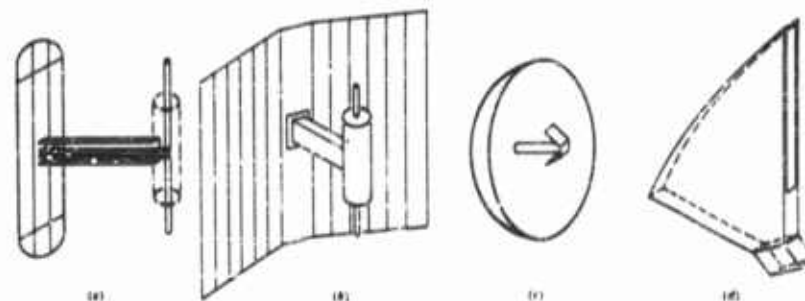
FIGURE 29-31 Reflector-type antennas. (a) Planar; (b) corner; (d) parabolic; (d) hoghorn.

ties can be predicted. The pattern and VSWR bandwidths are comparable with waveguide bandwidths. The pyramidal horn sketched in Figure 29-30a employs a tapered dielectric slug in a square-waveguide section to obtain circular polarization. The bandwidth of the linearly polarized horn of Figure

29-30b has been extended by the use of ridges in a manner analagous to that of ridged guides. Figures 29-30c and d shows a unique horn design* which provides approximately constant and equal E- and H-plane beamwidths, and a VSWR of less than 2:1 from 4 to 15 kMc. This structure uses "leaky" E-plane walls to achieve an approximately constant radiating aperture.

Reflector-Type Antennas. Relatively high-gain unidirectional patterns are a frequent requirement in ECM applications, particularly for ground-based or shipborne applications. A simple method for achieving this is to combine a wideband radiator with a planar, corner, or parabolic reflector. Figure 29-31 shows several examples of such antennas. Many of the radiators previously described are useful as feeds. The planar (a) and corner (b) reflectors are simple structurally and very satisfactory when relatively modest gains are required. The parabolic (c) reflector is usually more satisfactory when high gains are required. An "offset feed" arrangement with a parabolic reflector is particularly useful, since the feed is placed outside the main beam, and the poor side lobes and high VSWR's which can result from aperture blocking are thus avoided. The "hoghorn" structure of Figure 29-31d is unique among those shown, since it radiates a pattern which is broad in the plane perpendicular to the longer axis of the aperture and relatively narrow in the other.

All of the reflector antennas shown possess one property which is basic but usually undesirable; that is, the directive gain of these antennas depends on frequency, increasing approximately as f^2.

Surface-Wave Antennas. In airborne ECM applications, there is a frequent requirement for flush-mounted antennas with relatively high gain. Such requirements can often be met with the use of structures of the type sketched in Figure 29-32. These are so-called surface-wave antennas, whose gain re-
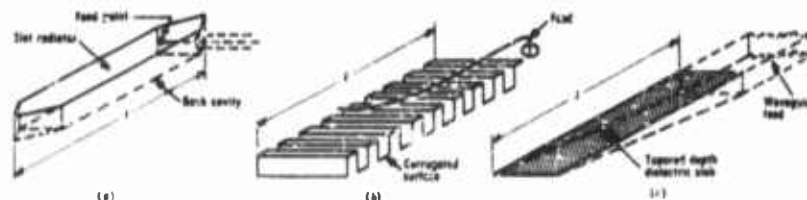
FIGURE 29-32 Surface-wave antennas. (a) TE antenna; (b) and (c) TM antennas.

sults from the guiding of the radiated wave by the structure. The patterns are endfire, and the gain is approximately given by

$$G = \frac{7l}{\lambda} \qquad (29\text{-}10)$$

*Developed at Melpar, Inc.

where l is the length of the array in the same units used for λ. The TE antenna sketched in Figure 29-32a produces a beam polarized parallel to the surface in which it is mounted, with the beam maximum tilted away from the surface by an amount inversely proportional to the antenna gain. The TM antennas of Figure 32b and c are polarized perpendicular to the mounting plane, with the beam maximum in that plane.

Antennas for Direction Finding. The principles of direction finding (D/F) for ECM are described in Chapter 10. A variety of antennas can be used as collectors. In D/F systems which utilize amplitude or phase comparison of the outputs of two or more fixed collectors, almost any of the antennas described previously can be used. One antenna which has been very successfully used for this purpose has not been described. This is the Luneberg lens with multiple feeds.* The basic configuration is sketched in Figure 29-33. The lens itself is in the form of a sphere of natural or artificial dielectric material, so arranged that the index of refraction varies with radius according to

$$ n = \sqrt{2 - \left(\frac{r}{R}\right)^2} \qquad (39\text{-}11) $$

where R is the outer radius of the lens. With such a lens, a feed placed at any point on its equator gives rise to a pencil beam emerging diametrically opposite.

Variable index media can be constructed in many ways. Among those used have been: low-dielectric-constant base materials loaded with high-dielectric-constant beads, high-dielectric-constant base materials loaded with holes, "path length" arrangements using parallel plates with variable spacing, and variable-density dielectric foams. One possible feed arrangement is sketched in Figure 29-33a. Here 15 independent linearly polarized feeds are spaced equally around the equator of the lens, with the axis of polarization skewed so that opposing feeds are cross polarized. This arrangement permits independent sampling of the 15 beams without serious interaction. Another arrangement is sketched in Figure 29-33b. This arrangement is intended for flush mounting on a plane surface. Since the lens action is two-dimensional in this case, the resulting beam is narrow in the plane of the lens but relatively broad in the plane perpendicular to the lens.

Several of the D/F schemes discussed in Chapter 10 employ rotating directional antennas. A number of difficulties arise in the design of practical collectors. There is, of course, the basic problem of achieving the useful gain

*Early work on this type of D/F collector was done at the Airborne Instruments Laboratory.
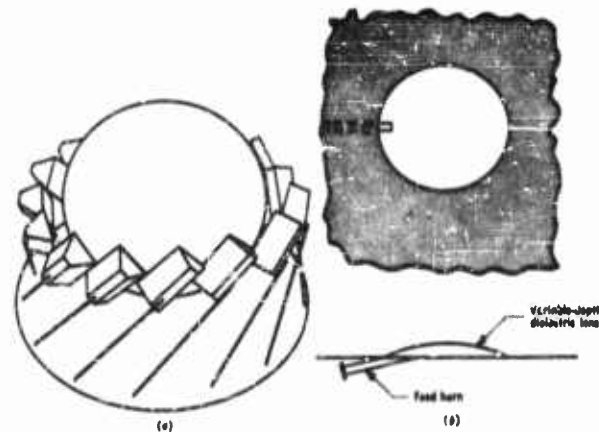
FIGURE 29-33   Luneberg lens. Two feed arrangements are shown.

and bandwidth in a limited volume. There is the problem of designing wide-band rotary joints. There is the major problem of distortion of the directional pattern resulting from "site effects," which is discussed briefly in Section 29-6.

Two types of rotary collectors have been used. In the first, the entire antenna system rotates (Figure 29-34a). In the other, sketched in Figure 29-34b, a circularly polarized directional antenna is fixed in such a position that
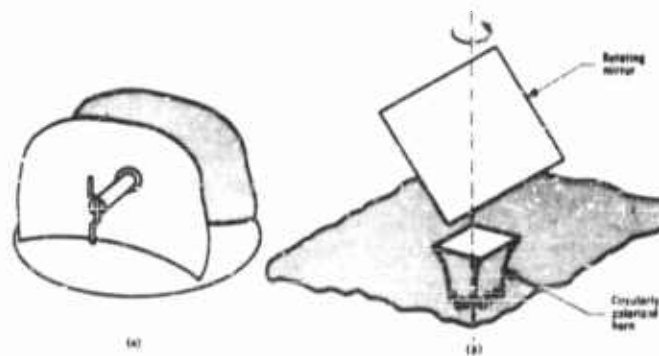
FIGURE 29-34   Two types of rotary collectors.

its pencil beam is pointed toward the zenith, and an inclined plane (sometimes shaped) mirror is arranged to rotate about the axis of the fixed antenna in such a fashion that the beam is caused to sweep in azimuth. The

obvious advantage of such an arrangement is the elimination of a rotary joint. The principal disadvantages are two: the polarization of the beam is essentially restricted to circular by the optics of the arrangement, and the gain (collecting area) of the mirror arrangement is inherently lower than would be the case for a conventional collector with the same aperture area.

Collectors in which the entire antenna rotates take several forms. Figures 29-34a and 29-35 show two examples in which feed/reflector combinations
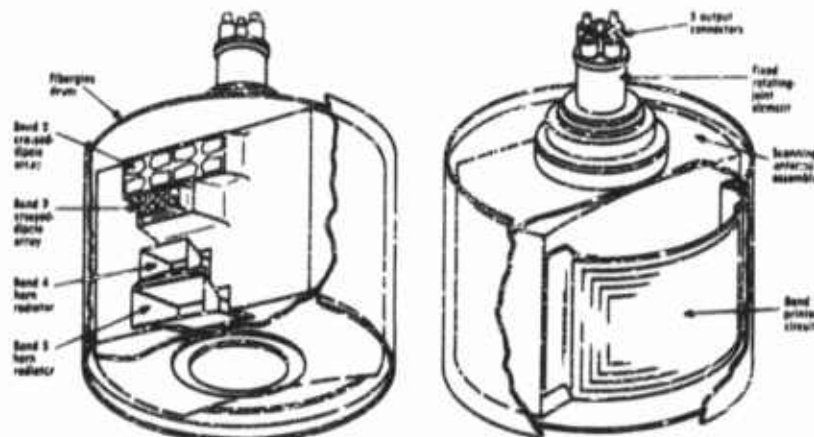
FIGURE 29-35 Five-band circularly polarized collector.

are used. The antenna of Figure 34a, a part of the APA-69, is useful over bandwidths of up to 5:1, but at best its electrical performance has been compromised in order to achieve a minimum size. Figure 29-35 shows a different approach, in which antennas for the five standard ECM bands from 550 to 10,750 Mc are combined in a single assembly, capable of rotation at high speeds.* A novel rotary joint providing for the five bands independently is incorporated in the design. Here the individual bands are covered by simple circularly polarized antennas. The performance obtained represents a compromise with the objective of minimum size.

29.6 Antenna Siting Problems

It is impossible to leave a discussion of ECM antennas without some consideration of the effects of siting on antenna performance. "Siting" effects cover a broad range of phenomena, all of which have to do with the distortion of the wavefront due to the presence of other bodies in the vicinity of

*The ALA-12 collector, developed at Dorne & Margolin, Inc.

the antenna. Included are reflections, refraction, diffraction, shadowing, and mutual coupling. (This list is not all-inclusive, and the categories mentioned are not free from overlapping.)

To deal with reflections first: It is obvious that any large metallic body will reflect radio waves; and this is equally true, but not equally obvious, for any large body whose intrinsic parameters differ from those of free space. (In the case of dielectric surfaces both reflected and transmitted components occur. The VSWR for reflected waves is equal to the square root of the ratio of the dielectric constants on the two sides of the interface.) The most common instance of reflection effects is the change in the free-space parameters of an antenna when it is located near the ground. These effects are well known. When the antenna is elevated above a plane earth, the reflected fields combine with the direct rays to create a lobe structure in the radiation pat-

FIGURE 29-37   Bearing error vs. true bearing, A/A-17 D/F collector.

FIGURE 29-36   Radiation patterns of a vertical stub UHF antenna located on the under-fuselage of a B-30 aircraft.
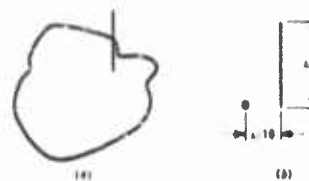
FIGURE 29-38   Diffraction pattern of dipole parallel to edge of flat plate.

terns. The higher the antenna (in wavelengths) the greater the number of lobes produced, and the higher the gain at the maximum of the lobe nearest the horizon.*

In operational installations of ECM antennas, the effects of reflections can be very severe. Two examples which indicate the severity of the problem are shown in Figures 29-36 and 29-37. Figure 29-36 shows the radiation pattern, in the three principal planes, of a $\lambda/4$ stub operated at 1000 Mc, the antenna being mounted on the bottom centerline of a B-50 aircraft. The deep lobing resulting from reflections, principally from the engine nacelles, is obvious. Figure 29-37 shows the measured bearing error of an APA-17 D/l^2 collector located on the bottom centerline of a RB-66 at a frequency of 450 Mc for vertical polarization. The very large bearing errors encountered are due to the distortion of the phase front of the incoming wave by reflections from various portions of the aircraft, principally the swept-back wings and the engine nacelles.

Refraction effects are usually associated with such meteorological phenomena as ducting and tropospheric scattering, and as such are discussed in Chapter 31.

Diffraction phenomena are encountered in every instance in which radiation extending into the region of the "shadow" of an obstacle is observed. Figure 29-38 shows the diffraction effects of the edge of a thin metal plate, $\lambda/4$ wide, level with and spaced $\lambda/10$ from a dipole (Figure 29-38b). All of the radiation appearing in the first quadrant of the polar plot (Figure 29-38a), as well as the distortion of the pattern evident in the sector opposite the diffracting edge, is due to diffraction. Generally, such diffraction fields are characterized by large phase rotations with small changes in angle. Diffraction also accounts for the back lobes in reflector type antennas.

Mutual coupling is a troublesome problem in many antenna installations, particularly where guy wires, feed cables, masts, and other structures of substantial length and relatively small diameter are located in the vicinity of the antenna. The mutual coupling which exists gives rise to parasitic radiation which combines with the direct field in such a way as to produce lobes and nulls or other pattern distortions. One example, special in nature but

*The increase in gain near the horizon gives rise to the practice of mounting antennas at the highest possible elevation, as on the masthead of a ship. This must be practiced with care, for at least two reasons. In the first place, at UHF frequencies the numerous nulls produced at elevation angles near the horizon may seriously degrade the system effectiveness. Second, in most installations, elevating the antenna increases the cable loss, since the location of the associated equipment is usually fixed. Where the usual propagation factors apply and the system parameters are otherwise fixed, the optimum height is that at which the cable loss is 1 neper, or 8.68 decibels. Above that height the signal loss due to cable attenuation increases faster than the gain due to increased height.
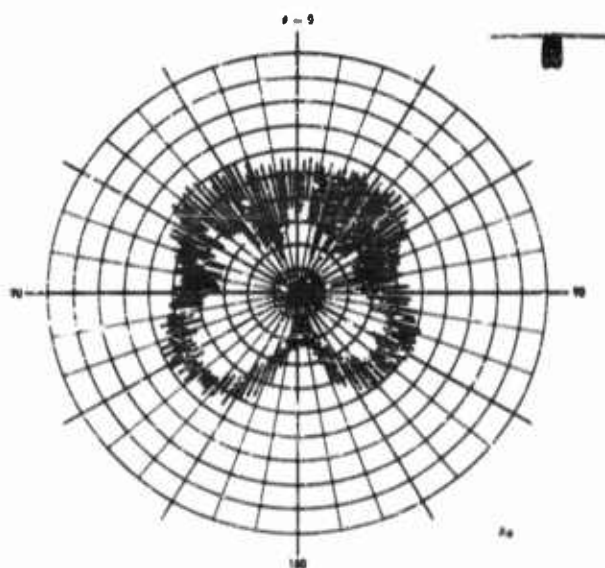
FIGURE 29-39   Helicopter rotor modulation of vertical monopole pattern.

Illustrating a problem common to many installations, is shown in Figure 29-39. This figure shows the distortion of the 30 Mc horizontal-plane radiation pattern of a 7-foot vertical whip antenna located on the tail boom of an H-19 helicopter by mutual coupling to the rotor blades. This pattern was taken by rotating the blades relatively rapidly while rotating the model rather slowly on a conventional antenna pattern range. The distortions of the radiation pattern is accompanied by equally gross variations in the input VSWR of the antenna.

REFERENCES

1. Jordan, E. C., *Electromagnetic Waves and Radiating Systems*, New York, Prentice-Hall, 1950.

2. Chu, L. J., "Physical Limitations of Omnidirectional Antenna," *J. Appl. Phys*, Vol. 19, p. 1165, December 1948.

3. Fano, R. M., "Theoretical Limitations on Broadband Matching of Arbitrary Impedances," *Tech. Rept. No. 41*, MIT Research Laboratory of Electronics.

4. Tanner, R. L., "Theoretical Limitations to Broadband Impedance Matching," *Electronics*, Vol. 24, February 1951.

5. Battelle, R. B., and H. E. Singhaus, "Airborne Jamming Antenna Study—Methods for the Evaluation of Antenna Patterns and the Derivation of Optimum Patterns," *Interim Rept.* 1, USAF Contract No. AF 33(606)-5584, Stanford Research Institute, May 1959.

6. Withrow, W. E., "Study of Power Requirements for X-Band Jamming from Surface Vessels, *Naval Research Laboratory Rept.* 4455, December 17, 1954 (CONFIDENTIAL).

30

Supplementary Circuits and Techniques

L. W. EVANS

30.1 Introduction

In this chapter an associated group of techniques is described which, when properly applied, will enhance the operation of countermeasure equipment and systems. The main subheadings of the chapter are Receiver Circuits, Analyzer Circuits, Transmitter Circuits, and Recording Techniques.

30.2 Receiver Circuits

30.2.1 Strong Signal Elimination

In countermeasure or electronic intelligence (ELINT) search procedures the signals of maximum interest are often very weak, demanding a system sensitivity which often pushes the state-of-the-art. In instances of this kind very strong signals emanating from nearby sources may mask the weaker signals of great interest. It is possible to eliminate strong signals by gating them out on an amplitude basis.
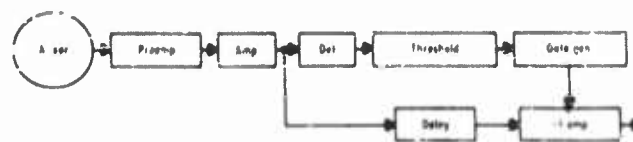
FIGURE 30-1 Block diagram of a strong signal elimination unit

The block diagram of Figure 30-1 illustrates a strong signal elimination unit. The design of the unit is straightforward. The unit arranges for strong signals to gate themselves out of the receiver video. The gating can be done before or after detection.

It should be noted that gating on the basis of amplitude will not eliminate the side effects of harmonic mixing. A signal of sufficient strength may cause the harmonics of the difference frequency from the mixer to be amplified in the i-f amplifier and appear as a weaker signal in the out᠎   . f the receiver at an erroneous frequency. The elimination of these sigt.   discussed in Section 30.2.2.

30.2.2 Rejection of Superheterodyne Receiver Responses Due to Mixer Harmonics

When very strong receiver input signals mix with the local oscillator (LO) signal in the nonlinear mixing element, harmonics of the difference frequency are generated along with the difference frequency. For example, a strong signal whose difference frequency is one-third the intermediate frequency will produce the third harmonic of the difference frequency at the intermediate frequency. This third harmonic will be amplified and appear at the receiver output as a weaker signal at the frequency to which the receiver is tuned. The undesirable effects of difference frequency harmonic generation are the reception of undesired signals, erroneous frequency information and possible interference with signals of interest.

The reception of the spurious signal which is the result of the difference frequency harmonic generation cannot be eliminated by the strong signal elimination technique described in the preceding section. This spurious reception occurs at different frequency settings. For example, if the intermediate frequency is 30 mc and a strong signal exists at 1000 mc, a strong receiver output will appear when the LO frequency is 1030 mc. But because of harmonic generation in the mixer, a signal may also appear when the LO frequency is 1015 mc, 1010 mc, etc. These spurious signals will vary widely in amplitude and can be gated out only after they have been recognized as spurious signals. The spurious response is recognized by the presence of a strong difference frequency which is a submultiple of 30 mc.

The block diagram of a harmonic rejection unit is shown in Figure 30-2. The mixer output is split into two channels. On the basis of a 30 mc intermediate frequency and an 10 mc bandwidth, the lowest order harmonic which would produce an output between 25 and 35 mc would be less than 17.5 mc. The next would be about 8 to 12 mc, etc. With the extension down
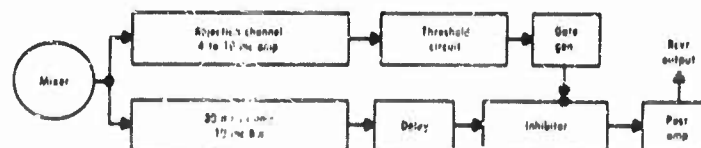
FIGURE 30-2   Block diagram of a rejection unit which will prevent spurious receiver responses due to mixer harmonics

to 4 mc it is likely that if there are any higher harmonics which will qualify, such as the 30th harmonic of a 1-mc difference frequency, there will also be harmonics of frequencies greater than 4 mc which can be detected in the rejection channel passband.

The delay in the straight through channel allows for the rise time in the gate that is generated in the rejection channel. Any signal appearing at the output of the mixer in the frequency range of 4 to 18 mc above a set threshold level causes gates to be generated which will prevent signals from appearing at the receiver output due to difference frequency harmonics. The threshold level is best set by experiment.

30.2.3 Local Oscillator Power Stabilizer

When voltage tuned local oscillators are used with scanning superheterodyne intercept receivers, widely varying LO power is often the result. The most undesirable effect of the variable LO power is that the receiver noise figure is seldom optimum. It is possible to place a variable attenuator in the LO power line and regulate the LO power for optimum noise figure by maintaining a constant mixer crystal current.

It should also be noted that the LO drive affects the conversion loss and thus changes the over-all receiver gain. The output impedance of the mixer varies inversely with the LO drive, and affects the bandpass shape of transitionally coupled i-f input circuits, which are often used where a broad i-f bandwidth is desired.

The attenuator pictured in Figure 30-3 is used to stabilize the LO power of a Carcinotron used in an intercept receiver. The attenuator consists of a short coaxial line. In the dielectric region of the line is a piece of ferrite. This ferrite normally introduces 12 db attenuation at microwave frequencies. However, when a strong transverse magnetic field is applied, the attenuation will drop to as low as 3 db. The attenuation is controlled by the current in
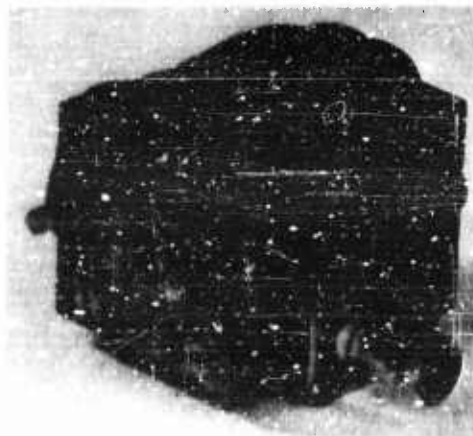
FIGURE 30-3   Variable attenuator for LO power stabilizer

the magnetizing coil. The attenuation can be varied over a 9 db range to maintain constant crystal current. The crystal current regulator is thus a closed loop servo. The servo speed is limited by the inductance of the magnetizing coils and the driving power of the current source. These characteristics are dictated by the scanning speed of the intercept receiver.

LO power stabilization can also be obtained by controlling the accelerating, or in some cases the grid, of backward-wave oscillators. Controlling these electrodes results in frequency pulling of about 1 percent.

30.2.4 An Image Suppression Receiver for Pulsed Signals

Superheterodyne receivers in which the mixer is continually accessible to all signals in the r-f band are subject to frequency ambiguity of the received signal. In such intercept receivers the unsuppressed image doubles the signal environment for the associated analysis and recording equipment. Standard subcarrier bands are listed in Table 30-I.

An image suppressing receiver is described capable of more than 60 db image rejection. The technique employed is a combination of outphasing and gating techniques. About 15 db image rejection is obtained by outphasing. This 15 db represents sufficient difference in signal levels so that gates may be generated to gate out strong image signals.

The block diagram of the image suppression receiver is shown in Figure 30-4. The r-f input is fed in phase through a power splitter and two directional couplers to two mixers. The LO power is fed 90 degrees out of phase to the two mixers. The i-f voltage output of channel 1 lags channel 2 output

TABLE 30-I. Standard Subcarrier Bands

Channel	Center Freq. (cps)	Lower Limit (cps)	Upper Limit (cps)	Deviation (%)	Freq. Response (cps)
1	400	370	430	±7.5	6
* 2	560	518	602	±7.5	8
* 3	730	675	785	±7.5	11
* 4	960	888	1032	±7.5	14
* 5	1300	1202	1398	±7.5	20
* 6	1700	1572	1828	±7.5	25
* 7	2300	2127	2473	±7.5	35
* 8	3000	2775	3225	±7.5	45
* 9	3900	3607	4193	±7.5	60
* 10	5400	4995	5805	±7.5	80
* 11	7350	6799	7901	±7.5	110
* 12	10,500	9712	11,288	±7.5	160
* 13	14,500	13,412	15,588	±7.5	220
* 14	22,000	20,350	23,650	±7.5	330
* 15	30,000	27,750	32,250	±7.5	450
16	40,000	37,000	43,000	±7.5	600
17	52,500	48,560	56,440	±7.5	790
18	70,000	64,750	75,250	±7.5	1,050

Optional Bands: A. This band may be employed by omitting the 30 kc band.
B. This band may be employed by omitting the 22 and 40 kc bands.
C. This band may be employed by omitting the 30 and 52.5 kc bands.
D. This band may be employed by omitting the 40 and 70 kc bands.
E. This band may be employed by omitting the 52.5 kc band.

A	22,000	18,700	25,300	±15	660
B	30,000	25,500	34,500	±15	900
C	40,000	34,000	46,000	±15	1,200
D	52,500	44,620	60,380	±15	1,600
E	70,000	59,500	80,500	±15	2,100

* Preferred.
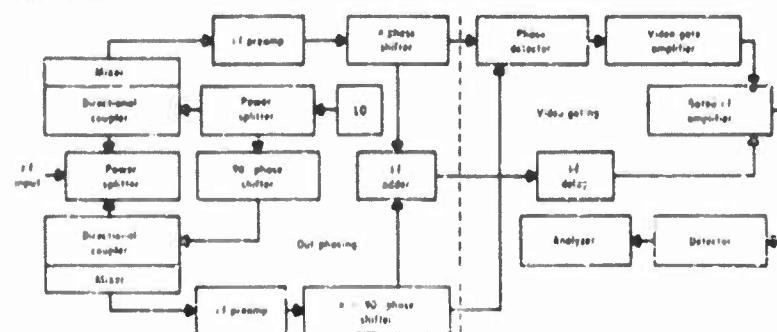
FIGURE 30-4 Block diagram of image suppressing receiver

by 90 degrees when the LO frequency is *higher* than the signal frequency. Channel 1 output leads channel 2 output by 90 degrees when the LO frequency is *lower* than the signal frequency. In the i-f circuitry of channel 2 the phase of the voltage is retarded relative to channel 1 voltage by 90 degrees. The voltages in the two channels are then added. The voltages reinforce when the LO frequency is higher than the signal frequency and cancel when the LO frequency is lower than the signal frequency. This is the basic outphasing method of image rejection. Because of practical limitations in producing the phase shifts and maintaining equal signal amplitudes at the adder inputs, the image will not completely cancel. The amount of rejection by the outphasing technique varies with errors in the phase shift circuits and amplitude balance. For example, if the inputs to the adder are matched within 1 db and the phase shifters are all within 10 degrees of 90 degress, the image rejection will be of the order of 20 db. This amount of rejection alone is insufficient for practical applications.

To increase the rejection of the image to 60 db the gating circuits of Figure 30-4 are added. At the input to the adder for the signal, two voltages nearly in phase are applied, and in the case of the image, two voltages nearly out of phase. The phase detector in the gating part of the receiver will pro-

FIGURE 30-5   Model of image suppression receiver

vide opposite polarity video outputs in the cases of signal and image. This output is amplified and used to gate off an i-f ampli. er when an image pulse is present. The video gate is not required until the ge is strong enough to overcome the basic rejection by outphasing. Hence, the signal-to-noise ratio at the phase detector is sufficient for the generation of a noise-free gate.

A model of an image receiver is shown in Figure 30-5.

30.2.5 Interference Reduction by Side Lobe Suppression

Interference arriving through antenna side lobes may seriously degrade the performance of detection equipment. A device designed to eliminate pulsed side lobe interference is illustrated in Figure 30-6.

The technique consists of adding to the normal intercept receiver an identical receiver called the interference receiver which is fed from the same local oscillator. The gain of the interference receiver is adjusted to be as close as possible to that of the normal receiver. The interference receiver is fed by an antenna whose gain characteristic is omnidirectional in the azimuth plane and closely approximates the search antenna in the vertical plane. The relationship of the two azimuthal channel gain functions is shown in Figure
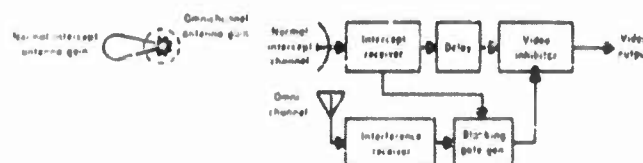
FIGURE 30-6 Block diagram of side lobe interference suppression system

30-6. The antenna gain of the interference channel is adjusted to be larger than that of any side lobe of the normal channel but smaller than the main lobe.

In the blanking gate generator, the outputs of both receivers are compared in amplitude. Should the signal in the auxiliary channel be larger than the signal in the intercept channel, a blanking gate is generated which prevents the signal from appearing at the output of the normal channel. This corresponds to the case of a side lobe signal since a side lobe signal will be larger at the interference receiver output than at the normal receiver output. If the signal in the normal channel is larger, this indicates the presence of a main lobe signal.

The delay is introduced to compensate for the inherent delay of the blanking gate. The circuit suffers a loss in sensitivity due to the blanking periods. This loss increases with the side lobe interference.
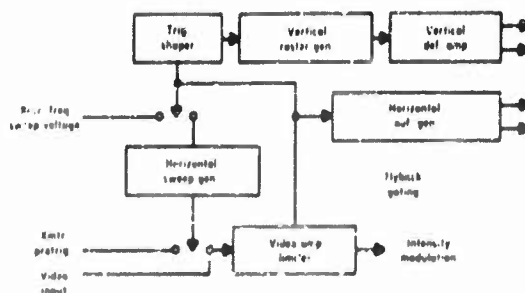
FIGURE 30-7   Block diagram of visual storage display unit

30.2.6 Panoramic and Video Display of the Visual Storage Type

A panoramic display utilizing a visual storage tube permits the use of very low receiver frequency scan speeds. In addition, an intercept history is presented for evaluation or recording.

To make optimum use of the storage capabilities of the visual storage tube, the electron beam is intensity modulated by the received video and a vertical raster is employed which increases the vertical deflection in steps at the end of each frequency scan. Indefinite storage can be obtained by the use of the Hughes Memotron storage tube. Shorter storage times from 30 seconds to a few minutes can be obtained with the use of gray scale Tonotron type tubes.

On the pan display an intercepted signal takes the form of a vertical line. Intermittent signals and frequency shifts are easily detected. Noise pulses appear as scattered random spots on the face of the display. Weak signals in a noise environment can be readily detected because of the integrating properties of the display.

Figure 30-7 shows a block diagram of a typical panoramic display using a visual storage tube and Figure 30-8 shows a photograph of such a display. The incoming video is applied to the grid of a dual control pentode which amplifies weak signals, limits strong signals, and provides a means of video blanking during flyback of the receiver sweep.

Because of the low sweep speeds encountered, the receiver frequency sweep voltage is d-c coupled to the horizontal deflection plates of the storage tube. If the receiver stops, the location of the spot on the storage tube retains its correct frequency relationship to the receiver sweep. For mechanically tuned receivers a triangular voltage can be generated by mechanically driving a potentiometer. Triggers for the raster generator can be generated by limit switches.

Figure 30-9a shows an example of the panoramic display. The raster steps up rapidly at the bottom of the display and slowly at the top. A strong
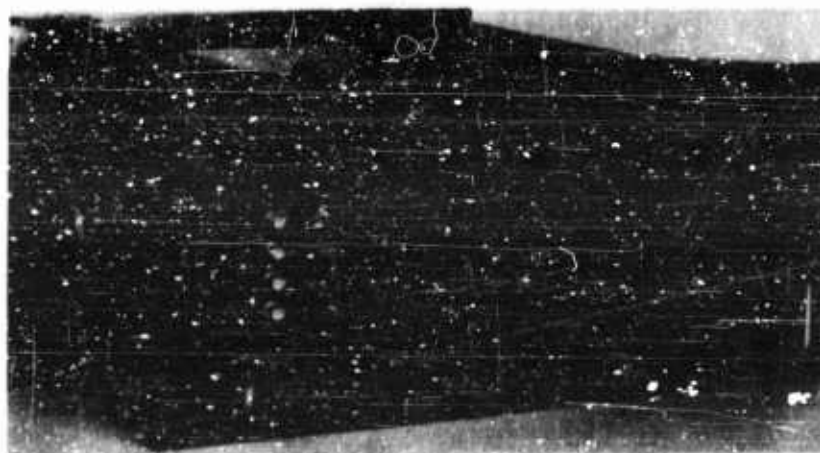
FIGURE 30-8   Visual storage display unit

signal and image intercept is shown in the center of the display with a shift
in frequency near the top. The other spotted areas are noise bursts. The dis-
play examples of Figure 30-9 are negatives of actual photographs.

The visual storage tube can also be used to display a history of low fre-
quency video data. The writing speeds are such that this type of display is
inadequate for wideband data.

If an internal sweep generator is used which is triggered by the incoming
video pulses, prf data will be displayed on the storage tube. Using the raster
scan, missing pulses or time synchronized signals are readily detected. Figure
30-9b shows a display where the horizontal deflection sweep is triggered by
the input video. The vertical raster is stepped up so that missing pulses in
the input video are readily detected.

The visual display can also be used to monitor the operation of the range
unit described in Section 30.3.11. In this case the display sweep generator is
triggered by a pretrigger from the jammer. Synchronized replies will store
as a line and are readily detected. Unsynchronized replies are stored as back-
ground random spots. This mode of operation is illustrated in Figure 30-9c.

The circuit that provides the vertical steps synchronized to the receiver
sweep is shown in Figure 30-10. This circuit provides the following desired
features:

1. Each step is synchronized to the receiver sweep and is independent of the
    receiver sweep speed.
2. The vertical step occurs only on flyback and is a true step function.

ELECTRONIC COUNTERMEASURES

Range (miles)

Sweep
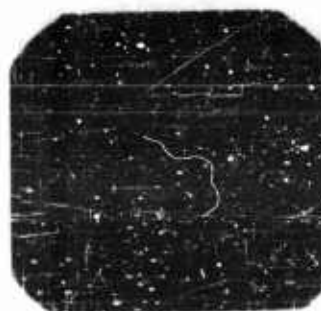
Time

Frequency
(A)

Range (miles)

Sweep
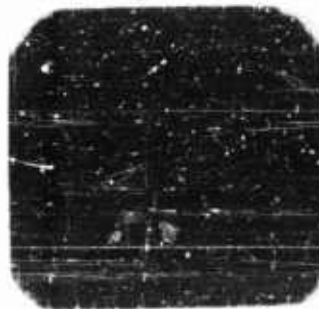
Time

Frequency
(B)

Range (miles)

Sweep

Time

Frequency
(C)

FIGURE 30-9   Applications of a visual storage display
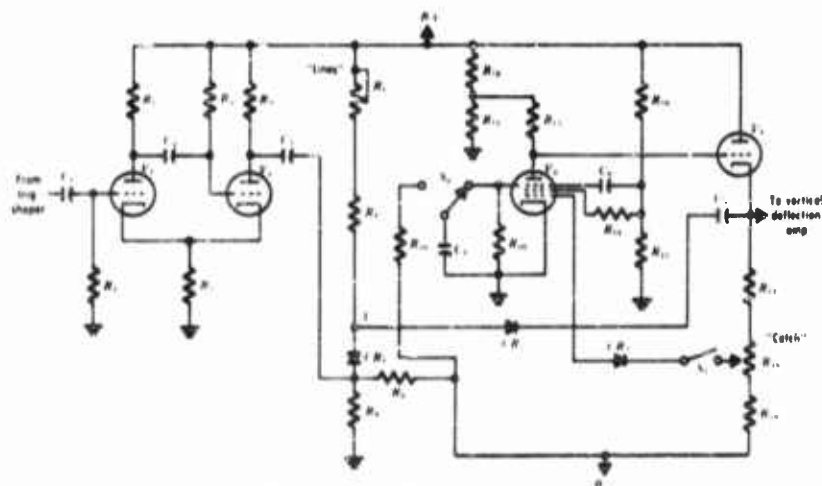
FIGURE 30-10 Raster generator

3. The magnitude of each step is controllable over a wide range and may be adjusted at any time by the operator.

4. The vertical excursion of the beam may be stopped at the top and held until the operator desires to reset it to the bottom or it may be made to recycle automatically. Erasure of the stored information may or may not be accomplished with recycling.

The circuit is basically a free running phantastron which is prevented from free running by means of R8, R9, CR1 and CR2. Point A of Figure 30-10 is held at a negative potential by the divider consisting of R8 and R9; this places a back bias on CR2 and the charge on C5 is trapped causing the phantastron to stop at any point in its cycle (except on flyback) the instant the negative voltage is applied.

As used in the panoramic display the negative voltage is applied at all times and the phantastron is gated to its free running state by the application of a positive pulse at point B of Figure 30-10. This positive pulse places a reverse bias on CR1 and removes the reverse bias from CR2 and the phantastron is permitted to free run for the duration of the pulse. The actual voltage change at the output of the circuit is determined by the width of the gating pulse and by the time constant of C5, R6, and R7. If the input pulse is held at a constant width, then the magnitude of the step is determined by R6 and may be changed at any time.

The positive pulse for gating the phantastron is generated by a monostable

flip-flop which in turn is triggered by the flyback of the receiver sweep. By making R2 and R7 preset adjustments the minimum and maximum number of "lines" in the vertical direction can be set.

Since the actual stepping in the vertical direction is determined by the positive pulsewidth, it is possible to use the negative pulse present at the other plate of the monostable flip-flop to gate off any incoming video during the actual stepping action.

By the addition of diode CR3 (and closing S1) the phantastron can be prevented from recycling even though the positive input gating pulse continues to be applied. At some point on the cathode return resistor of V4 there is a potential that equals the grid voltage of V3 just prior to flyback of the phantastron. Prior to this time the grid voltage of V3 is rising and the cathode voltage of V4 is falling and CR3 has a reverse bias applied to it. At the instant that the voltages on the cathode and anode of CR3 become equal the action of the phantastron is stopped. The point of its cycle at which these voltages become equal is determined by the setting of R18 and this "catch" control can be set so that the action is stopped at the top of its cycle by closing S1 and adjusting R18. With S1 open the vertical deflection of the beam is in an "automatic" recycle condition. With S1 closed the mode is termed "manual" recycle. The manual recycling is accomplished by the application of a negative pulse to the suppressor of V3. Switch S2 serves this purpose; it is a momentary contact type which discharges the negatively charged condenser C4 through R20.

The flyback of the phantastron is in a direction opposite to the individual steps and can be used to automatically erase the visual storage tube. By means of a switch this function can be made automatic or manual.

This flyback voltage can be used to actuate a framing camera such as the KD2 and an output of the camera used to erase the information stored on the storage tube. In this way a complete record of all intercepts can be maintained. This mode of operation is discussed in detail in Section 30.5.1.

30.2.7 Frequency Transfer

It is often necessary to transfer frequency from the local oscillator of an intercept receiver to the local oscillator of a passive track receiver. This may be accomplished using many different techniques. The particular method described here allows for storage of the LO frequency in a resonant cavity thus providing optimum use of the intercept equipment.

The frequency transfer system is described with the aid of the block diagram of Figure 30-11. A small amount of r-f energy from the intercept receiver local oscillator is coupled into a resonant cavity. The cavity is tuned to the LO frequency by a motor driven slug. The tuning of the cavity is ac-
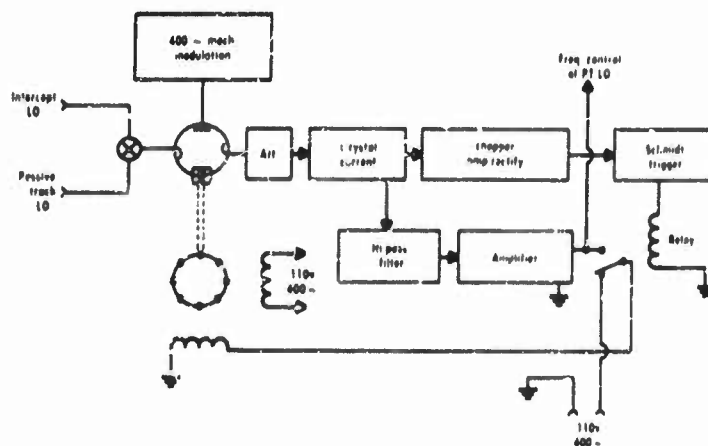
FIGURE 30-11   Block diagram of a frequency transfer system

complished in two steps. 1) In slew, the cavity is considered to be tuned off. Power is applied to the motor to bring the cavity near the proper frequency. At the point where crystal current appears, sufficient gain in the slew channel will operate the relay which connects one winding of the motor to a servo amplifier. 2) In the servo position the tuning of the cavity is nulled by the servo. The cavity is made to serve as a microwave discriminator by mechanically deflecting one side of the cavity at an audio frequency. This generates an error signal for the servo.

Once the cavity is nulled with the intercept LO frequency, the cavity motor is turned off and the frequency stored. The frequency can be transferred to any other LO at any time. To accomplish this transfer another servo similar to the one described here is required to null the LO of the passive track receiver.

The frequency transfer system described here has been greatly simplified. The particular application will require specific compensating networks to stabilize the servo. Using Carcinotron local oscillators in S-band it is possible to transfer frequencies with a precision of 0.5 mc and better. The transfer time can be made as small as 100 milliseconds.

Another frequency transfer system is suggested by Melchor and Vartanian, (*Proc. IRE*, February 1956). In this system the cavity described above is replaced by a coax line filled with the paramagnetic substance, hydraxyl. A magnet is placed around the coax line and provided with two windings. A d-c current through one winding determines the frequency of the sharp resonance line of hydrazyl. The a-c winding provides a means of modulating the r-f fre-

quency. A means of slewing and nulling the d-c current can be provided in a manner similar to that demonstrated above.

30.2.8 Voltage-Frequency Linearization of Backward Wave Oscillators

Applications of a microwave receiver using a backward wave oscillator as a voltage tuned local oscillator are complicated by the nonlinear relationship between the collector voltage and LO frequency. A typical tuning curve is
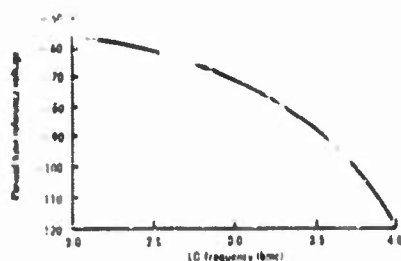
Figure 30-12 Typical tuning curve for a backward wave oscillator

illustrated in Figure 30-12. It is usually desirable to have the LO scan in frequency linearly with time. This can be accomplished by distorting a linear sweep voltage before it is applied to the collector. The distorting network is called a linearizer.

A linearizer can be constructed using the on-off characteristics of diodes as nonlinear circuit elements. In this way a desired function is replaced by a number of straight line segments. Such a circuit functions as a variable attenuator with the diodes switching in or out arbitrary amounts of attenuation. The voltage level at which a particular diode switches is fixed by connecting the diode to an appropriate bias voltage. A linearizer circuit is illustrated in Figure 30-13a. Its input and output characteristics are shown in Figure 30-13b.

A complete BWO frequency sweep circuit using a linearizer is shown in Figure 30-14. A low level linear sweep voltage is applied to this circuit. The input voltage is distorted, shifted in level, and used as a reference voltage in a high voltage power supply regulator to obtain the voltage level required by the BWO collector.

30.3 Analyzer Circuits

30.3.1 The Automatic Threshold Circuit

The automatic threshold circuit is useful in regenerating the pulse output of intercept receivers. It will provide almost noise-free video from the receiver to analysis or recording equipment over wide ranges of receiver gain and noise level.

The automatic threshold circuit establishes a clipping level just above the
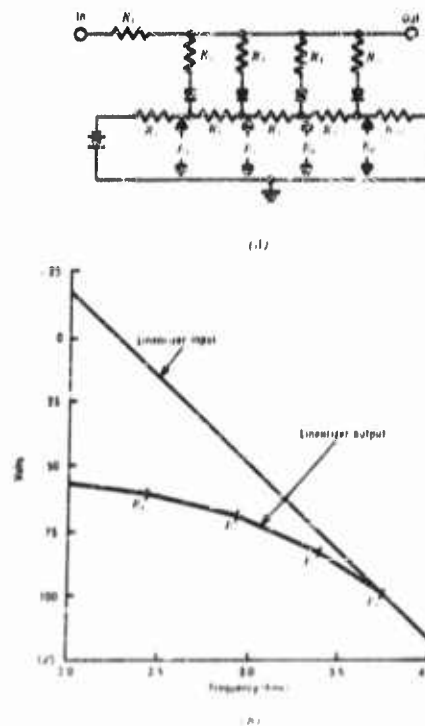
FIGURE 30-15   A linearizer circuit

noise and regenerates a 0.2 volt slice of the receiver video at this level. The level is automatically adjusted by the noise level. This is accomplished by amplifying and detecting the noise.

The block diagram of the automatic threshold circuit is illustrated in Figure 30-15. The noise amplifier has a low frequency cut-off such that it has negligible response to the video. The diode slicer clips the receiver video just above the noise level. The 0.2 volt slice is taken to prevent overdriving the video amplifier which follows. A threshold circuit is shown in Figure 30-16.

The threshold circuit provides video for recording and analysis in a manner insensitive to receiver gain setting or noise level. The signal level required for the regeneration of the receiver video is nominally 8 db above that signal level which yields a S + N to N ratio of 2:1.

30.3.2 Pulse Extractor

Pulse extractors are useful in extracting unwanted pulse trains from the
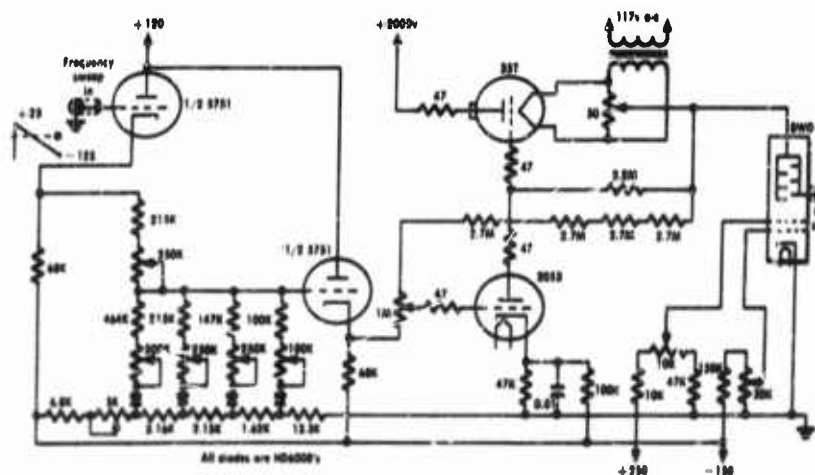
FIGURE 30-14   BWO sweep circuit using a linearizer
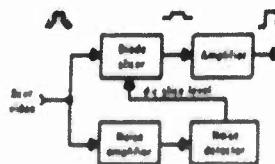
FIGURE 30-15   Block diagram of
automatic threshold circuit

intercept receiver output. The prf of the unwanted signal must be manually set into the unit. Or, it may be made to scan in prf, lock on the first intercepted prf, and extract the pulse train to which it is synchronized.

The pulse extractor can also be used as a signal selection unit and generate a signal to stop the frequency scan of a receiver when it has intercepted a signal with the preset prf. In this mode of operation the extractor has a harmonic susceptibility and should be used only at high repetition rates when the probability of a harmonic stop is remote.

The synchronizing circuit in the pulse extractor is a triggered delay circuit of instant recovery. The delay will synchronize with a pulse train whose period is slightly greater than the delay the first time it is triggered by a

Figure 30-16   Schematic of automatic threshold circuit

pulse in that train. The delay will not lose sync as long as the pulse train is present unless another train appears which better qualifies in prf. In this case the synchronizing circuit will switch pulse trains.

The delay circuit of instant recovery is made up of two delays in cascade with an inhibitor which prevents the first from being triggered before the delay of the second is complete. The two delays provide time for recovery of each individual delay circuit. The delay circuit is illustrated in Figure 30-17.

Figure 30-17   Delay circuit of fast
recovery

The block diagram of the complete pulse extractor is shown in Figure 30-18. The Gate, G, at the output of the rapid recovery delay circuit is used to determine when the instant recovery delay is synchronized with a pulse train. If this gate is filled with video pulses several times in succession, it can be concluded that the extractor is synchronized with a pulse train and this same gate may be used to extract pulses from the input video. This gate determines the resolution of the unit as well as the amount of video extracted. The gate may have any width desirable. However, a gate of 5 to 10 percent of the delay is nominal.

FIGURE 30-18   Pulse extractor block diagram

The "And 2" circuit is used to determine when Gate G is filled with video pulses. The output of the "And 2" circuit triggers a multivibrator which gates a step counter. A predetermined count of 4 or 5 will operate the Schmidt trigger and at this point pulse extraction begins. The step counter is dumped when the Gate, G, is void of pulses twice in succession. This is accomplished by the parts of the circuit represented by I_5, D_5, R_1, and "And 3." The reader is left with the simple task of determining the operation of this part of the logic.

30.3.3 Signal Selection Circuits

Signal selection circuits are used in detecting the presence of signals known to be threats on the basis of signal parameters. When the presence of a signal threat is detected, an alarm is generated and a jammer can be brought into action. The threats may be identified on the basis of any combination of pulsewidth, prf, and the number of pulses in the pulse code groups. The signal selectors can be connected in cascade to set up the proper combination of parameters. In the cascade connection it is logical that the order of pulse-width, prf, pulse code be maintained because a pulsewidth discriminator is least susceptible to interference and a pulse code presence detector is the most susceptible to interference.

30.3.4 Pulsewidth Discriminator

The pulses from an intercept receiver can be examined on an individual basis. Thus the pulse width discriminator has no significant interference problem. However, to obtain rectangular pulses from an intercept receiver, the receiver must be wide band. This means a sacrifice of sensitivity and an invitation for interference which may cause difficulty for other equipments. Therefore, it is profitable to compromise and permit a limited degradation of pulse shape for greater frequency selectivity and receiver sensitivity.

The pulse width discriminator will be of less value now and it is dangerous to make the channel widths too narrow or have too little overlap. The block diagram of a single channel of a channelized pulse width discriminator is
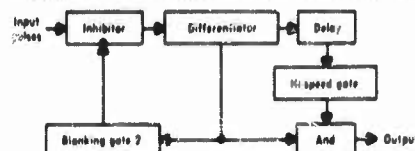
FIGURE 30-19 Block diagram of pulse width discriminator channel

shown in Figure 30-19. The input pulses are differentiated. The leading edge is delayed an amount equal to the minimum width of pulse which will qualify. The delayed trigger generates a gate equal to the acceptance range of the channel. If the trailing edge of the input pulse falls within this gate, the channel will provide a short pulse output. This pulse can be used to operate a prf filter or a pulse code selection unit. The inhibitor and Gate 2 are used to prohibit the qualification of a pulse code group as a long pulse.

30.3.5 PRF Selector

The prf selector recognizes the presence of a pulsed signal with a preset prf in the output of an intercept receiver. The selector described here will perform its function when interfering pulse trains are present. The prf selector is a filter which responds to the fundamental frequency component of a pulse train. Special techniques are required to eliminate subharmonic ambiguities because pulse trains are rich in harmonic content. The mathematical analysis below shows that the harmonic ambiguities can be eliminated on an amplitude basis providing the input video pulses are standardized.

The operation of the prf filter is described with the aid of the block dia gram

FIGURE 30-20 Block diagram of prf filter

frequency components whose amplitudes are given by[*]

$$A_K = 2E_0 \frac{S}{\tau_r} \left[\frac{\sin \pi K f_r S}{\pi K f_r S} \right]$$

where

E_0 = the pulse amplitude
S = the pulsewidth
τ_r = the pulse period
f_r = repetition frequency
K = order of the harmonic

[*]Principles of Radar 2nd Ed. Staff MIT Radar School pp. 4-14.

The examination of the above expression reveals that the value inside the brackets is a sin x/x function and is equal to unity for the first several harmonics. Therefore

$$A_K \sim \frac{2E_0 S}{\tau_r}$$

Hence, for the harmonics of greatest concern, the amplitudes of the frequency components are inversely proportional to the *fundamental* pulse period. For example the 1000 cycle component of a 500 cycle pulse train is one-half the 1000 cycle component of a 1000 cycle pulse train. Therefore, the subharmonic ambiguities can be eliminated with an amplitude selector. The pulse standardizer of Figure 30-20 must regenerate every pulse in each pulse train and only one pulse for a pulse code group or echo train. This requirement is imposed by the amplitude selection circuit. The filter will normally be low Q. A notch filter or a pulsed Hartely oscillator will provide a prf resolution of about 3 percent. As the bandwidth is broadened it becomes more difficult to discriminate against harmonics of low prf signals. It is difficult to design the prf filter so that it can be manually tuned. It is a laborious tracking design problem. Complete prf analysis with miniaturized filters can be accomplished with about 12 filters per octave. A simple prf filter circuit is shown in Figure 30-21.

FIGURE 30-21   PRF filter

The prf filter will operate satisfactorily when pulse trains are intermixed. It is possible to generate gates which will gate out the video pulse that has caused the filter to ring. Thus the video signal that has qualified in prf can be gated to a pulse code selector. The gate which gates the video can be generated from the tip of the sinewave generated by the filter. The gates

generated in this manner are in the order of 25 percent of the period. Shorter gates can be generated by cascading the prf filter and extractor of Section 30.3.2 with the filter and pulse extractor set to the same prf.

30.3.6 Pulse Code Presence Detector

Interleaved pulse trains place pulses close together at intervals such that a simple pulse code presence detector will generate many false stops as a result of interference. A computer of considerable magnitude is required with a pulse code selector to measure and handle the data necessary for perfect operation of the selector unit. On the other hand, it seems that the most practical approach is to depend upon frequency, pulsewidth, and prf selection to remove the interference and construct the simplest pulse code selector possible.
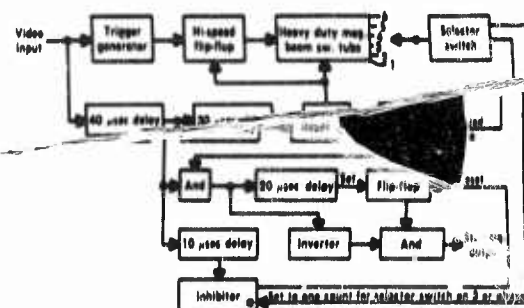
FIGURE 30-22   Block diagram of pulse code presence detector

The block diagram of Figure 30-22 illustrates a pulse code presence detector with a limited interference capability. With the selector set on 3 or above, the selector will produce a receiver scan stop signal after two successive proper counts of pulses in the input gate of 40 microseconds. In the "successive" counting, counts of 1 are ignored. This minimizes the effect of interference of a single pulse train. However, with the selector set on 2 or less, an interfering pulse train will usually generate a false stop or prevent a legitimate stop depending on circumstances. The manner in which this logic is accomplished in the pulse code presence detector illustrated in Figure 30-22 is left to the reader.

30.3.7 New Signal Qualification

The new signal qualification unit provides an alarm for an ELINT procedure when a new signal appears. The new signal qualification described here is on the basis of frequency only, although it is possible to make use of other parameters by extending the techniques.

The method of detecting new signals uses two Radechon storage tubes

In a scan to scan comparison system which compares the frequencies of all signals intercepted on one scan of a receiver with the frequencies of the signals intercepted during the previous several scans as controlled by the partial erasure of the storage tubes. Any signal appearing suddenly at a new frequency qualifies as a new signal and generates an alarm and a new signal stop.

The intercept receiver provides a horizontal deflection voltage to the storage tubes proportional to frequency. One tube is in the "read" condition and the other in a "write" state. The receiver video is standardized and used to intensity modulate both tubes together. If an intercepted signal is an "old" signal, readout is obtained from the tube in the "read" state. This disqualifies the signal as a new signal. If no readout is produced, the circuits recognize this as a new signal, an alarm is generated and the receiver is stopped on frequency. Normally, when no new signal is found, at the end of each frequency scan the read and write functions are reversed for the next frequency scan.

The storage tubes consist of a standard cathode ray gun with electrostatic deflection. The storage area consists of a barrier grid placed in contact with a dielectric layer; a backing plate is in contact with the back side of the dielectric.

Storage is accomplished by applying a positive voltage to the backing electrode and turning the beam on. Readout is accomplished by returning the backing electrode to ground potential, positioning the beam to the desired point, and turning it on.

In this system an r-f type of readout is used. No r-f is applied to the tube in the write condition when a spot is stored. A 15 mc r-f voltage is applied to the vertical deflection plates of the tube in the read condition. The reading beam intersects the spot at a 30 mc rate and the barrier grid output is amplified by a 30 mc amplifier.

A block diagram is shown in Figure 30-23. The function is as follows:

The push-pull deflection amplifiers provide horizontal deflection on the storage tubes.

Flyback of the receiver is differentiated by the trigger amplifier, a blocking oscillator circuit. The blocking oscillator is used to trigger the storage switching circuit, consisting of a bistable multivibrator and cathode followers to furnish the read-write switching and gating voltages.

The receiver video is passed through a threshold unit to provide noise free video. The threshold video is applied to a gate trigger amplifier. The purpose of the gate is to prevent storing or reading out during flyback or when the receiver is in a stop condition. These two gates are mixed in an OR circuit and applied to the TRIG AMP/GATE.
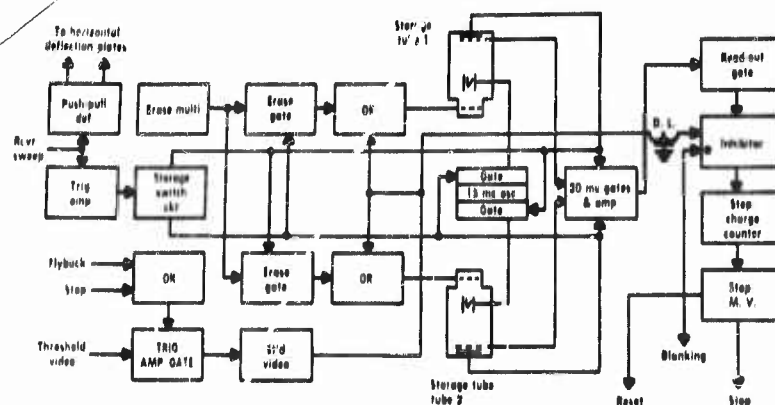
FIGURE 30-13 Block diagram of new signal qualification unit

The output of the TRIG AMP/GATE is used to trigger a monostable multivibrator thus standardizing the input video in amplitude and width.

Readout is gated through from the tube in the read condition and amplified by a 30 mc amplifier and detected. A trigger is generated from readout; this triggers a multivibrator producing a negative gate to inhibit the delayed standard video.

Should no readout exist this inhibitor passes delayed standard video and is counted in a step-charge counter. Consecutive pulses cause the counter to trigger the stop multivibrator. The purpose of the step-charge counter is to prevent random noise pulses from stopping the receiver and generating a false alarm.

After a stop is produced the stop bistable multivibrator is manually reset; at the same time blanking is supplied to the delayed standard video as the receiver leaves the passband, preventing stops from being generated on the same signal.

The 15 mc oscillator and gates provide the r-f read voltage to the appropriate tube.

To prevent saturation of storage a means of partial erase is provided by the erase generator. This consists of an astable multivibrator and gate circuits for applying erase to the tube in the read condition. Reading is an erase process. Amplitude and and prf are variable which allows adjustment of the erase time.

The erase voltage and standard video are mixed in an OR cathode follower circuit; this drives the grid of the storage tubes providing the intensity modulation needed for storage and readout.

30.3.8 PRF Correlators

In this system an r-f type of readout is used. No r-f is applied to the tube transmitter determine the effects of its jamming upon a victim beacon. When the jamming is effective the normal response of the beacon to its legitimate interrogati n will be altered. This change in the beacon output can be detected by observing the output of a passive track receiver tuned to the beacon frequency.

The change in the beacon response will take one of two forms. 1) the jamming may cause the beacon to fail to respond to its legitimate interrogation consistently, or 2) the jamming may cause the beacon response to contain extra pulses as it responds to the jamming transmitter. The equipments employed to determine when these effects are present are referred to as prf correlators.

The simplest way to determine whether the jamming is effective is to measure the beacon prf and detect any change in prf due to jamming. Another method employs a prf filter to detect the presence of the transmitter prf in the beacon response. Two correlators will be described here: 1) the Non-Periodicity Correlator which determines whether there are any missing or extra pulses in the beacon response and, 2) the Pulse Position Correlator which searches the beacon response for pulses in synchronism with the transmitter jamming pulses.

30.3.9 The Non-Periodicity Correlator

The Non-Periodicity Correlator examines the beacon response for missing or extra pulses. Two outputs are provided from the correlator. One output is generated when greater than a preset portion of the beacon pulses are missing. The second output is generated when greater than a preset percent of the jamming pulses elicit extra beacon responses.

The operation of the Non-Periodicity Correlator is described with the aid of the block diagram of Figure 30-24. The basic circuit is a prf circuit of fast response. The prf is measured by generating sawteeth between pulses. These sawteeth are peak detected for a measure of pulse period. The pulse period is integrated for a short time average and storage. The measuring and storing is a continuing process.

The individual sawteeth are compared with the short time average pulse period. When a pulse is missing in the beacon response the length of the sweep is doubled. A missing pulse is detected when a sweep is generated which is greater than the average pulse period.

When the passive track receiver is intentionally blanked to protect the mixer circuits from high power transmitters, beacon pulses are inadvertently missed. This means that a small percentage of the beacon pulses will be

FIGURE 30-24   Block diagram of Non-Periodicity Correlator
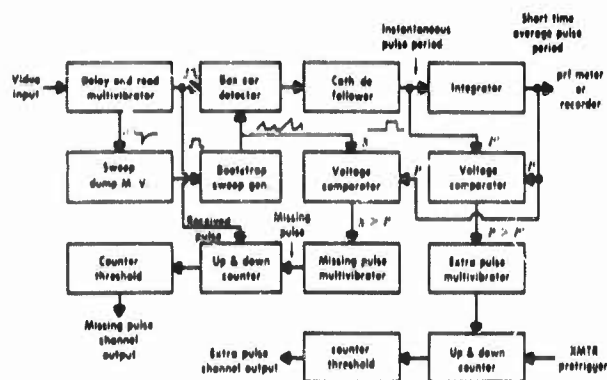
missed because of receiver-transmitter proximity. The up-and-down counter imposes the requirement that a certain percentage of the beacon responses must be missed before the pulse-missing channel provides an output. The counter counts up 7 for a missing pulse and down 1 for every received pulse. An output will be generated when more than 1 of every 7 pulses are missing. The selection of the ratio of 1:7 and the threshold level is arbitrary.

The presence of extra pulses in the beacon response is detected by comparing the instantaneous value of the peak detected sawteeth (instantaneous pulse period) with the short time average pulse period. An extra pulse will shorten the sweep and cause an abrupt decrease in the measured pulse period. When a certain percentage of the transmitter pulses result in extra pulses in the beacon response as determined by the second up-and-down counter, an output will be generated in the extra pulse channel.

30.3.10 Pulse Position Correlator

The Pulse Position Correlator is used to determine when the response of the victim beacon contains pulses in synchronism with the transmitter pulses. The beacon pulses are stored in their proper time phase position with respect to the transmitter pulses. When several beacon pulses fall in the same time position, with respect to the transmitter pulse, in succession or alternate succession a synchronized beacon output is detected and an output signal generated.

Several means are available for the storage mechanism. Time can be quantized and the incoming beacon pulses can be fed into a shift register. Another way of storing the pulse position data is to employ storage tubes such as the RCA Radechon cathode ray storage tube. The sweep is triggered

by the transmitter pretrigger and the tubes are intensity modulated by the
beacon pulses. A scan to scan comparison technique is used to determine
when the beacon pulses have a component in synchronism with the trans-
mitter pulses. A block diagram of the Pulse Position Correlator is shown in
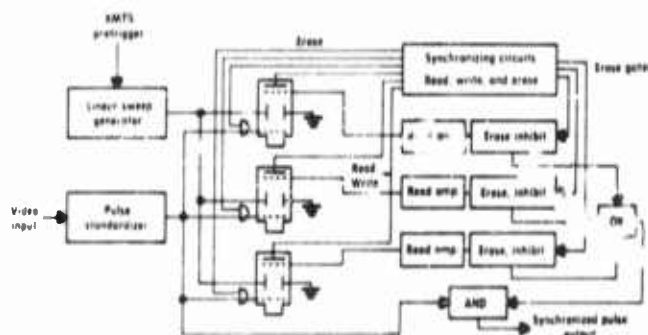
FIGURE 30-25   Block diagram of Pulse Position Correlator

Figure 30-25). The synchronizing circuits controls the programming of the
read, write, and erase functions of the tubes. When storing or writing, the
tubes in the "read" position are examined for readout. If any readout occurs
from the tubes, an output signal is generated. Three storage tubes are used
to facilitate the read, write, and erase function switching. Usually it is desir-
able to erase after each time the tube has been in the read condition. How-
ever, this requires that, to correlate, the beacon must respond to the trans-
mitter twice in succession. This may prove to be unlikely, particularly when
a high prf is used. Thus it may not be desirable to erase after every read
sweep. The beacon pulses can be stored longer by erasing less often, thereby
increasing the reliability of the correlator.

30.3.11 A Range Unit

One of the variables required to determine missile trajectory is range. A
range system is described which will obtain range information in the presence
of interference. The range data will be in error an amount corresponding to
missile beacon response time. The range unit employs a scan comparison
method with two Radechon storage tubes.

The range unit is illustrated in the block diagram of Figure 30-26. The
transmitter prf used is assumed low enough to eliminate the possibility of a
range ambiguity. When the transmitter is keyed for ranging, a sweep is
applied to each of two Radechon storage tubes The received video is stand-
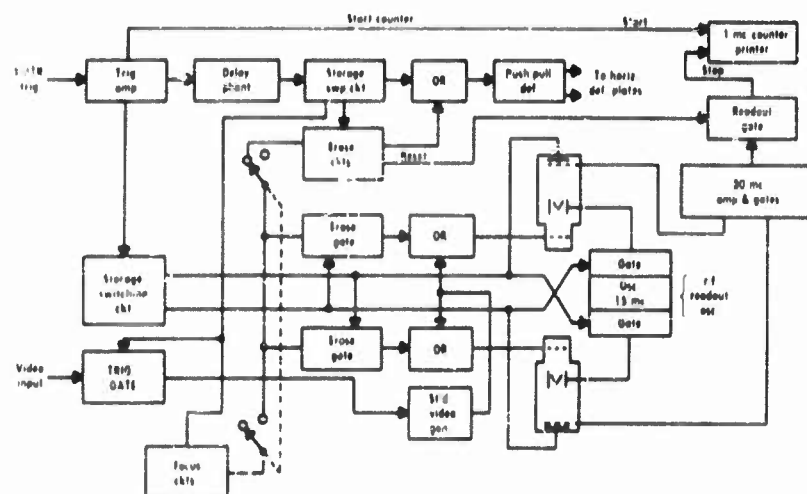
FIGURE 30-26 Block diagram of Range Unit

ardized and used to intensity modulate the tubes together. One of the storage tubes is in the "write" state; the other is in the "read" state. When a pulse is received it is stored in the proper time position with respect to the transmitter pretrigger on the tube in the "write" state. If at this point, any readout is obtained from the tube in the "read" state, it is assumed that these are replies from the transmitter pulses. In this regard we have essentially a pulse position correlator operating at a low prf. The range data is best obtained by starting a 1-mc counter with the transmitter pretrigger and stopping the counter with the correlated output from the storage and scan comparison circuits. Improved operation of the range unit can be obtained by eliminating the legitimate beacon responses with a pulse extractor.

For storage and readout operation refer to the New Signal Qualification Unit (Section 30.3.7).

The operation of the range unit is described as follows: A trigger amplifier is used to start a 1 mc counter and a delay phantastron, and also trigger the storage switching circuit from "read" to "write" condition.

The delay phantastron provides a fixed delay time to eliminate all signals within a minimum undesired range. This delay triggers a storage sweep circuit which furnishes sweep voltage for storage and also a gate voltage for allowing video to be stored.

The storage sweep circuit triggers an erase circuit on flyback which furnishes a second sweep for the tubes as well as an erase voltage applied to the

tube in the read condition by the erase gates, reading being an erase process. The tube in the read state is erased at the end of the storage sweep. The storage sweep and erase sweep are mixed and a push-pull deflection amplifier feeds the horizontal deflection plates of the Radechons.

The video is applied to a trigger gate circuit which allows a trigger to be passed only during the storage sweep. This triggers a monostable multivibrator, standardizing the input video. This standard video is then applied to the grid of the storage tubes through an OR circuit where it is mixed with the erase intensity gates.

Readout is accomplished by supplying a 15 mc r-f voltage to the tube in the read condition. The stored spot is intersected at a 30 mc rate and a 30 mc amplifier provides a readout pulse.

The readout is shaped and used to stop the 1 mc counter.

If no readout is produced, a reset is obtained from the flyback of the erase sweep which also stops the counter.

Focusing of the storage tubes is accomplished by storing in one pulse and reading out continuously with the focus multivibrator. The readout waveshape time is proportional to the spot diameter and can be adjusted by the focus control.

30.3.12 Signal Simulators

In the development, maintenance and test of complex intercept or active countermeasure systems, signal simulation is a prime requirement. It is important that maintenance personnel have test equipment which will enable them to do the simple tasks such as checking receiver sensitivity and also to accomplish tracker and transmitter programming. An engineering field test by airborne equipment should not be equired to check out a countermeasure system.

It is usually a simple task to construct a simulator tailored to the specific system requirements. The simulator can trigger or modulate a standard signal generator whose output can be coupled through directional couplers to the receiving and analysis equipment. Dry runs and system checkout can be accomplished by merely throwing a switch. Programming procedures can be simulated either on a video or r-f basis. It is often desirable to build simulation equipments to check out equipment subsystems to avoid involving the whole countermeasure system.

Examples of countermeasure simulators are listed below.

1. Scanning radar burst simulator; see Figure 30-27. The block diagram is self-explanatory.

2. Beacon signal and ranging simulator, simulating jamming effect of missing pulses and extra pulses. The beacon signal simulator of Figure 30-28 is

useful in quickly checking the Non-Periodicity Correlator, the Pulse Position Correlator and Range Unit. The simulator provides at one output a simulated beacon response where pulses may or may not be missing in variable amounts. The pulses are gated out in a random fashion. The output for the Pulse Position Correlator contains an unsynchronized beacon pulse train with mixed pulses at a controlled range which are synchronized with the simulated transmitter pretrigger. The transmitter pretrigger and Pulse Position Correlator outputs provide a means of checking range systems. The motor-driven delay simulates range rates.
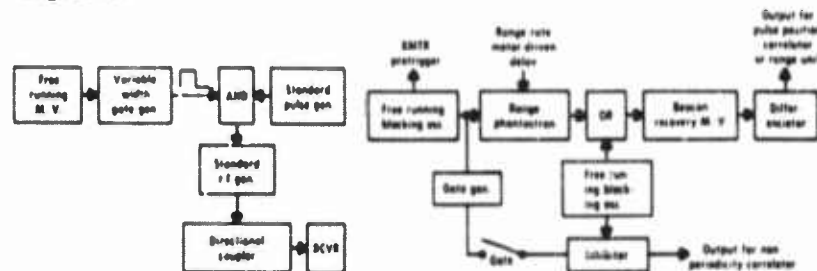
FIGURE 30-27 Scanning radar burst
simulator

FIGURE 30-28 Block diagram of beacon signal
and ranging simulator

30.4 Transmitter Circuits

30.4.1 Pulsed Transmitter Duty Cycle Monitor

The duty cycle of a pulsed transmitter must normally be controlled to prevent damaging the power oscillator tube. This problem requires special attention when variable pulsewidths and repetition rates are employed. Normally the pulsewidth is determined by the tactical situation and the prf is adjusted for the desired duty cycle. In most cases the prf is maintained at its maximum possible value depending upon the duty cycle specified by the power oscillator tube manufacturer.

The duty cycle of a transmitter is given by

$$D = Wf \qquad (30\text{-}1)$$

where D = the duty cycle
   W = the transmitted pulsewidth
   f = the repetition rate

The above relationship indicates that for a constant duty cycle the prf must vary inversely with the pulsewidth.

The basic repetition rate generator is a plate-to-grid coupled astable multi-vibrator with the grid return voltage E_u connected to a variable voltage source which provides a voltage proportional to the pulsewidth. The multivibrator is illustrated in Figure 30-29.
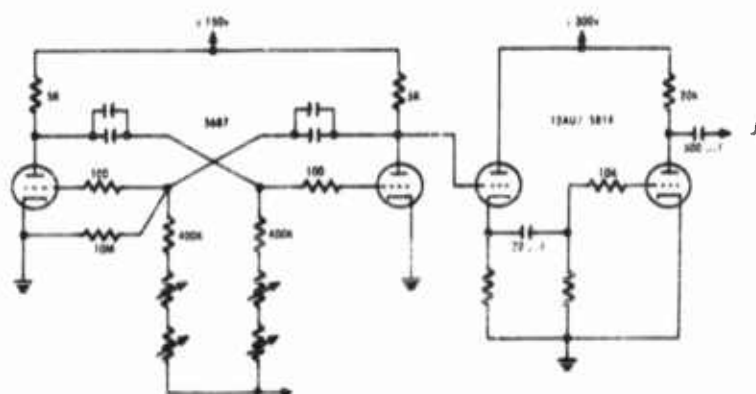
FIGURE 30-29   PRF multivibrator and shaping circuit

For a symmetrical multivibrator, the period of oscillation is given by

$$T = 2RC \ln \left(\frac{E_u + E_i}{E_u + E_c} \right) \tag{30-2}$$

where
E_u = grid return voltage
E_i = plate voltage swing during transition
E_c = tube cut-off voltage
R = grid resistance
C = plate-to-grid timing capacitor

Inverting we get:

$$f = \frac{1}{2RC} \ln \left(\frac{E_u + E_c}{E_u + E_i} \right) \tag{30-3}$$

For values of $E_u \gtrsim 0$

$$f = A + BE_u \tag{30-4}$$

The constants A and B can be evaluated graphically from Eq (30-3). The fact that the prf has a linear relationship to the grid return voltage makes it possible to use this circuit to provide a constant duty cycle for variable pulsewidths.

For the purpose of this discussion it is assumed that the pulsewidth of the transmitter is step-variable. By closing switches the width is increased by fixed increments. Figure 30-30a shows a bridge voltage divider circuit to provide voltage E_q as a function of the number of pulsewidth increments, N. In Figure 30-30b the circuit has been reduced to its approximate equivalent circuit. E_1 and E_q are voltages to be determined. K is a constant and n is the number of pulsewidth switches closed. It is assumed that the on-off switches controlling the number of pulsewidth increments switch the bridge resistors in and out.

From Figure 30-30,
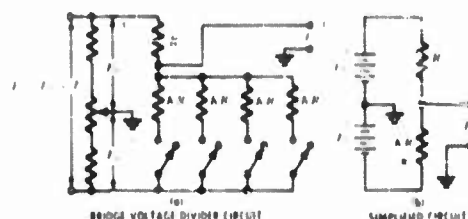
FIGURE 30-30   Bridge voltage divider

$$E_n = \frac{KE_1 - nE_q}{n + K} \tag{30-5}$$

Substituting E_q equal to A/B and combining Eq (30-4) and (30-5) yields

$$I = K \frac{(A + BE_1)}{n + K} \tag{30-6}$$

Multiplying each side of (30-6) by N we get

$$NI = NK \frac{(A + BE_1)}{n + K} \tag{30-7}$$

Note that N and n are different. N is the actual number of pulsewidth increments. n is the number of bridge switches closed. With no switches closed let N be 2. Then if K is equal to 2, Eq (30-7) reduces to

$$NI = K \ (A + BE_1) \tag{30-8}$$

For a constant duty cycle

$$K (A + BE_1) = \text{Constant} \qquad (30\text{-}9)$$

For small values of K such as 2, E_1 may be large. With E_1 of 450 volts it is possible to obtain a 10 to 1 variation in pulsewidth and control the duty factor within ± 10 percent.

The automatic control of duty cycle with continuously variable pulsewidth is possible by converting pulsewidth measurements into an analog voltage and using the same prf control circuit illustrated in Figure 30-29.

30.4.2 Transmitter Automatic Frequency Control

High performance countermeasure systems require that the jamming transmitter be capable of scanning and stopping in frequency rapidly. Also they must be accurately set on frequency at critical times. This is difficult to
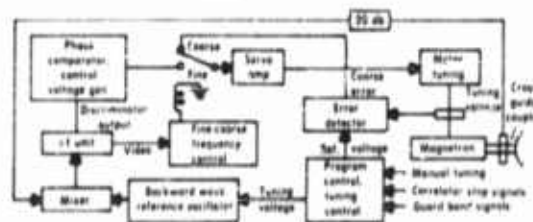
Figure 30-31   Block diagram of transmitter afc system
for mechanically tuned oscillators

accomplish when the transmitter is mechanically tuned. Two afc systems are described below. The first is applicable to mechanically tuned magnetron oscillators. The second is applicable to a voltage tuned carcinotron oscillator.

30.4.3   AFC System for a Mechanically Tuned Transmitter

A block diagram of an afc system for a mechanically tuned transmitter is illustrated in Figure 30-31. The frequency control is accomplished in two steps, coarse and fine. The coarse frequency control is a position servo which will set the transmitter frequency within 10 mc. At this point the fine frequency control takes over and fine tuning is accomplished. The backward wave reference oscillator is tuned with a d-c voltage. A potentiometer on the tuning shaft of the magnetron has a voltage applied. A position servo matches these two voltages to accomplish the coarse tuning.

For fine tuning of the transmitter, a small amount of r-f energy from the transmitter is mixed with r-f from the reference oscillator. The output of

the mixer is amplified in a 30 mc, wideband, i-f amplifier. The reference oscillator is normally 30 mc below the transmitter frequency. The image is removed by the coarse frequency control system.

The i-f amplifier provides an output of video pulses and also a discrim inator output. The discriminator output is pulses which vary in amplitude from −10 to +10 volts. The presence of the video pulses causes the afc system to switch from coarse to fine control. The pulses from the discriminator are stretched and peak detected to obtain a d-c control voltage to drive the fine tuning servo.

When a stop signal is applied to the afc circuit the reference oscillator is first stopped because it is voltage tuned. Then the transmitter will align itself with the reference frequency.

30.4.4 AFC System for Voltage Tuned Transmitters

The block diagram of Figure 30-32 illustrates an afc system by which a
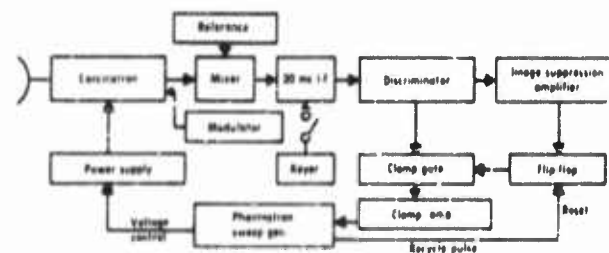
FIGURE 30-32   Block diagram of afc system for voltage tuned transmitter

high power backward wave carcinotron oscillator is controlled by a low level reference signal. The operation is fully automatic in the search and lock-on phases of operation. From any off-frequency condition the time required to lock on is less than one second. The system can track at rates of 300 mc per second per second with an error of 1 mc, and can be operated with any type transmitter modulation.

The afc system of Figure 30-32 has two modes of operation, search and lock-on. The search operation is useful to place the transmitter oscillator on frequency. The phantastron sweep generator generates a sawtooth of a 1-second period. This sawtooth controls the tuning voltage of the carcinotron by adjusting the reference voltage of the power supply regulator. During the search phase the phantastron runs free and cycles the transmitter tuning over and over.

The lock-on phase of the afc operation holds the oscillator on the desired frequency once it has been reached. To accomplish lock-on, a sample of r-f
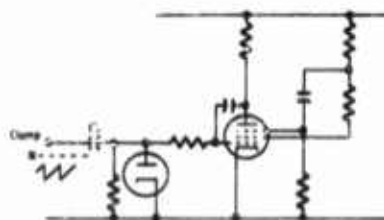
FIGURE 30-33 Frequency control phantastron

energy from the transmitter oscillator is mixed with the reference r-f in a broadband mixer. The output of the mixer is 30 mc which is amplified, limited, and detected by discriminattor. For all types of modulation the output of the discriminator is pulses. For all types of modulation except pulse, a keyer is used to pulse gate the i-f. The pulses are amplified and generate a proper waveform to maintain the phantastron in a quasi-stable condition in which the sweep voltage generated is stopped and held. In this condition the control loop is complete and the transmitter frequency will track the reference frequency. The phantastron circuit is shown in Figure 30-33.

The image suppression amplifier prevents locking on the lower sideband. The system uses the lower sideband to open the gate which will then allow pulses to reach the clamp circuits. This is illustrated in Figure 30-34. The operation is as follows: The transmitter starts sweeping at A and the first output from the discriminator appears at B. These are positive pulses which are not passed to the clamp ampli-

FIGURE 30-34 Lock-On operation of the BWM afc system

fiers because the gate is closed. At C pulses come from the sideband amplifier and open the gate. Positive pulses are no longer available from the discriminator and no pulses will appear at the clamp amplifier. The system is now susceptible to locking action when positive pulses are available from the discriminator. This happens at D when the discriminator output is first negative and then positive. If the reference pulses are lost for any reason, the gate is closed by the recycle action of the phantastron which resets the bistable flip-flop

30.4.5 Transmitter Guard Band Circuits

Guard band circuits are required to eliminate transmitter radiation at specific prohibited frequency bands. This is accomplished in the mechanically tuned and voltage tuned transmitters by blanking the modulator when the oscillator is tuned to these frequencies.

For the case of the mechanically tuned transmitter, guard band cams

operate in conjunction with roller activated microswitches to generate modulation blanking gates whenever the transmitter is tuned through the preset reference position.

In the case of the voltage tuned transmitter described in Section 30.4.1, it is convenient to arrange for the reference oscillator to be voltage tuned. Amplitude selection circuits are used in connection with the tuning voltage to generate modulator blanking gates whenever the voltage has specific levels. The same system can be used with the mechanically tuned transmitter by doing the same thing with the reference oscillator when it is voltage tuned.

30.5 Recording Systems

30.5.1 Continuous Recording of Intercept History

In the operation of a surveillance type receiving system it is often important to record both the time and the r-f frequency of each intercept. Slow speed chart records of receiver agc, frequency, and a timing signal have been used in the past with some success. A new method has been devised which provides considerable improvement in resolution of time and frequency data and also reduces the "bulk" of data required for a given period of time. This method has been successfully used with both mechanically scanned receivers and electronically scanned receivers.

This method uses a visual storage tube (Memotron) and an oscilloscope recording camera (type KD-2). A d-c analog of the receiver tuning is used for the horizontal deflection of the storage tube and a raster generator displaces each trace vertically. A noise-riding slicer operating on the receiver output develops a stretched, noise-free signal which is used to intensity modulate the storage tube beam. Thus, a blank screen indicates no signal, an intensity spot indicates a signal present, and the r-f frequency is shown by the horizontal position. A vertical line indicates a signal present each time the receiver scans that frequency. Depending upon the number of traces in the raster and the receiver sweep speed, the vertical displacement of the spot indicates the time of the intercept.

The raster generator used in this unit allows selection of from 40 to 400 lines per raster. Compensating circuits allow a single knob control for this function and regardless of the number of lines per raster, the entire usable scope face is utilized. The number of lines per raster are not limited to 400. Rasters with several thousand lines are feasible since the number of lines are determined by the resolution required in time of intercept.

The camera is triggered on the recycling of the raster generator, a single frame exposure is made just prior to erasure, and a new raster commences immediately. Each picture also shows the data panel of the camera which includes a clock and a sweep second hand. If receiver sweep speed can be held

constant, the time of intercept of a signal can be resolved from the photograph with a resolution of approximately ±0.005 of the time required to generate one raster.

The sweep speed used by a scanning type receiver in a surveillance operation is usually dictated by the intercept probability requirement. One factor in determining the speed is the number of pulses it is necessary to receive to record an intercept. In this method of recording, a single received pulse is sufficient to verify an intercept. This feature will usually allow the receiver
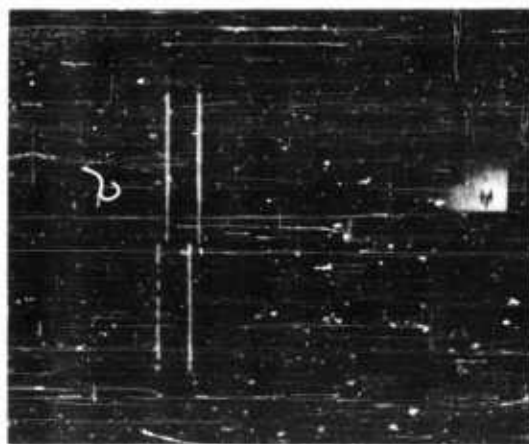
FIGURE 30-35 Typical frame-by-frame recording of data
recorded by Memotron storage display and KD-2 camera

to be tuned faster thus improving the intercept probability. Figure 30-35 is an example of data recorded by the Memotron storage display and KD-2 camera.

30.5.2 Fast Continuous Film Recording

For wideband data recording it is possible to use a high speed camera which passes film at a high rate of speed by the face of a cathode ray tube. The motion of the film provides the time scale. The video to be recorded is applied as vertical deflection of the cathode ray tube beam. This type of recording consumes film at a high rate but this can be expected of wideband recording systems.

A Warrick camera uses film speeds of 125 feet per second and can stop

and start rapidly without breaking the film. Appropriate data analysis equipment can magnify the film data to 60 microseconds per inch.

30.5.3 Multiplex Recording Systems

FM-FM telemetry techniques are readily applied to multiplex recording systems. In general, standard FM telemetry packages will produce a composite signal which can be recorded on a single 100 cps to 100 kc track of magnetic tape. The tape bandwidth is divided between the multiplexed signals.

A block diagram of a multiplex recording system is illustrated in Figure

30-36. The data to be recorded are applied to voltage controlled oscillators (VCO) and cause the oscillator to shift over a given frequency band. A table of RDB standard subcarrier bands is shown as Table 30-1. The output of the VCO's are mixed and the composite signal is recorded directly on a magnetic tape recorder.

Figure 30-36 Block diagram of the three channel multiplex recording system

A timing signal can be recorded on one of the channels. On play-back, a set of filters are used to separate the several data channels. These filters are called demultiplexing equipment.

The multiplex recording system is, in general, very costly and complex. The calibration problem is also multiplied. It would probably be preferable to use a 14 channel tape recorder rather than multiplex 14 channels onto one channel of magnetic tape.

A multiplex recording system will not be required unless the number of required channels is greater than the number of recording tracks available on a single tape. Then the use of a multiplex system simplifies the time correlation problems of data analysis.

30.5.4 Time Base for Recording Mediums

There are many recording mediums: magnetic tape, chart recorders, strip film continuous motion, frame by frame photography, digital printers, etc. When records are made on more than one instrument during a single operation, time correlation between the various recorders is often difficult especially when remote stations are involved. The time correlation problem is such that a universal time standard is necessary; and time must be recorded along with the data on every record, often with accuracies of a fraction of a second.

Radio station WWV provides the required universal time reference. This is a time reference only. Precision timing equipments capable of generating accurate timing information for long periods of time must be used to generate the actual timing signals. These timing signal generators can be referenced to WWV at convenient intervals.

Timing signal generators are commercially available. The one described here is manufactured by the EECO of Los Angeles. The precision timing generator will drift ±1 second per month. If the unit is referenced to WWV frequently, timing information is available for high speed records within one millisecond of absolute time. The timing code is a modified binary form applied as amplitude modulation to a 100 cycle sinusoid for high or low speed records. The code is illustrated in Figure 30-37.
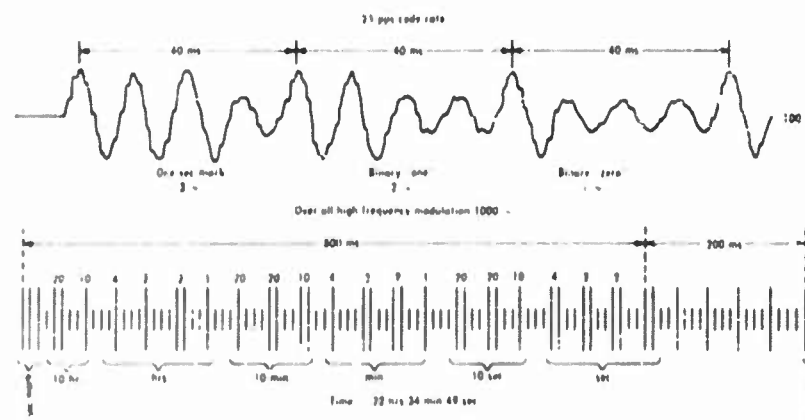
FIGURE 30-37   Time code

BIBLIOGRAPHY

1. Brick and Galejs, "Radar Interference and Its Reduction," *Sylvania Technologist,* July 1958. (UNCLASSIFIED)

2. Evans, L. W., "Separation of Interfering Pulse Trains on the Basis of PRF," EDL M69. Sylvania Electronic Defense Laboratories, Mountain View, Calif. (CONFIDENTIAL)

3. Evans, L. W., "Non-Periodicity Correlator," EDL M91. (CONFIDENTIAL)

4. Greener, A., "Image Suppressing Receiver for Pulsed Signals," EDL M115. Sylvania Electronic Defense Laboratories, Mountain View, Calif. (CONFIDENTIAL)

5. Harper, J. W., "Pulse Width Discriminator," EDL M98. (CONFI   TIAL)

6. Heenacker, L., "Voltage-Frequency Linearization of Backware   ve Oscillators," EDL M88. Sylvania Electronic Defense Laboratories, Mountain   w, Calif. (UNCLASSIFIED)

7. MacKenzle, M., "High Power C   inotron AFC," The University of Michigan Countermeasure Symposium Proceedings 1956. (UNCLASSIFIED) portion of SECRET publication)

8. Smith, J. I., "Pulse Code Presence Detector," EDL M124. Sylvania Electronic Defense Laboratories, Mountain View, Calif. (CONFIDENTIAL)

31

Propagation

A. T. WATERMAN, JR.

31.1 Introduction

Radio-wave propagation has an inherent role in nearly all applications of electronic countermeasures. Whether one is receiving radiations from an enemy equipment or emitting radiations that will affect an enemy equipment, the radiations must be propagated through the intervening space. Normally that space is occupied by the earth's atmosphere, and it is the role played by the various atmospheric constituents that makes propagation a dynamic and at times critical consideration.

This chapter will aim at surveying those aspects of propagation which are of most direct concern in countermeasures. With this viewpoint, it will not attempt to be a comprehensive treatise on propagation as such. Rather it will concentrate on factors which directly affect such quantities as signal level, angular deviations, and bandwidth limitations. Because of this special emphasis, the topics discussed will be introduced in a manner that is not entirely systematic from a propagation viewpoint. Effects which have a relatively simple influence on transmitted signals will be introduced first, even though they may in some cases involve a complicated physical process. Correspondingly, other effects involving simpler physical processes may be deferred to a later discussion if their role in countermeasures is a confusing or unpredictable one.

No deliberate attempt will be made to cover the various frequency ranges systematically. However, the above mentioned treatment will in some respects

involve a sequence that begins at the high frequency end of the spectrum and proceeds downward.

There will be a need to limit the discussion in many instances. Electronic countermeasures, in the broad sense, includes many forms of activity, extending over a vast frequency range. To delineate the coverage more clearly, emphasis will be placed on those frequency ranges and circumstances in which radar countermeasures are important. This is not to imply that only radar countermeasures are being considered. Indeed, specific reference will be made at times to countermeasures against communications and other radiating systems. However, such areas as infrared, visible, and ultraviolet radiation will not be discussed. Similarly the VLF end of the spectrum will be given only light treatment. The ranges covered will, for the most part, extend from a little below 30 megacycles to a little above 100 kilomegacycles.

Finally, the discussion will be aimed at present-day and immediate-future needs. No attempt will be made to speculate on the long-range outlook. Such problems as appropriate countermeasures for an interspaceship communications link will hinge on equipment not yet developed, frequencies not yet chosen, and properties of the interplanetary medium not yet firmly ascertained. The propagation environment discussed here will be largely that in which the earth's atmosphere plays the dominating role: propagation between ground, air, and satellite. For convenience in anticipating the discussion to follow, an outline of the material to be covered is given here.

Discussions of line-of-sight propagation and transhorizon propagation constitute the remainder of this chapter. Under line-of-sight propagation are considered free-space propagation, molecular and ionic absorption, tropospheric and ionospheric dispersion and refraction, and polarization. Under transhorizon propagation are considered tropospheric and ionospheric refraction and scatter, propagation via meteor trails and via aurora, diffraction, artificial modifications of the propagation medium, and transhorizon ranges. The list of ECM categories below will be used as a guide in referring to specific applications, although this list may oversimplify the subject of countermeasures and the role propagation plays in it. As each type of propagation is discussed, its potential role in this classification of countermeasures activities will be mentioned.

Active countermeasures
    Power jamming
    Spoofing (deceptive jamming)
Passive countermeasures
    Detection
    Locating
    Detailed signal analysis

31.2 Line-of-Sight Propagation

A large portion of the transmissions which are of concern in military applications deal with propagation between two points lying within line-of-sight of each other. Although these situations are often less interesting than those involving more extended coverage, they are of basic importance and serve as a convenient general reference.

31.2.1 Free-Space Propagation

If radio power P_T is transmitted from an antenna having gain G_T (with respect to a theoretical isotropic radiator), then the power P_R delivered to the matched load of a receiver connected to an antenna of gain G_R at distance d from the transmitter is

$$P_R = P_T \cdot G_R G_T \cdot \left[\frac{\lambda^2}{16\pi^2} \cdot \frac{1}{d^2} \right] \qquad (31\text{-}1)$$

where λ is the wavelength (measured in the same units as d) (References 1, 2, 3, and 4). The quantity in square brackets is a geometrical quantity only. It, or rather its inverse, represents the loss which must be overcome by the other quantities, which are all man-made (transmitter power, antenna gains, and receiver sensitivity). When expressed in terms of frequency f in megacycles and distance d in miles the inverse of the quantity in square brackets is

$$4.56 \times 10^3 f^2 d^2 \qquad (31\text{-}2)$$

It is useful to have a feeling for some typical magnitudes of this free-space transmission loss (Reference 5). For example, at a frequency of 10,000 mc and a distance of 100 miles, this loss expressed in decibels is 157 db. Such a transmission loss can be overcome, for example, by a transmitter power of 100 kw (80 dbm), a receiver sensitivity of -40 dbm (crystal video) and a combined antenna gain (transmitting plus receiving) of 40 db. All this adds up to 160 db, giving a 3-db margin over the free-space loss.

Because transmission loss varies as the square of both frequency and distance in Eq (31-2), the sample figure of 157 db loss would apply also to a frequency of 1000 mc and 1000 miles, or to a frequency of 100 mc and 10,000 miles, or alternatively to 100,000 mc and 10 miles. Thus, other things being equal, free-space propagation favors the lower frequencies. To the extent that other things (transmitter power, receiver sensitivity, antenna gains) are not equal, the lower frequencies tend to be even more strongly favored.

The above sample figures indicate that a typical high-powered radar at X-band could be detected at distances of the order of 100 miles without excessively sensitive receiving equipment. Another 20 db increase of receiver sensitivity (to −60 dbm) would extend this range to around 1000 miles. On the other hand, at higher frequencies, good receiver sensitivities and high transmitter powers are harder to come by, so that transmission ranges are limited by free-space attenuation alone. There is thus a transition region in the neighborhood of 10 to 30 kmc below which a high-powered ground-based transmission may be received anywhere within line-of-sight in the earth's atmosphere. The reciprocal situation—ground-based reception of airborne transmission—is also, of course, true.

These remarks apply to free-space transmission only. Other factors may enhance or diminish received signal strengths. These factors are discussed in the next sections.

31.2.2 Line-of-Sight Propagation—Absorption

The simplest modification of line of-sight propagation introduced by the atmosphere is the literal absorption of power by certain atmospheric constituents (References 1 and 3). All of the absorption mechanisms are frequency sensitive, some extremely so. Broadly speaking, there are two general categories of absorption operative in the frequency ranges of interest. The first is molecular absorption, principally by water vapor and oxygen. It is effective in varying degrees at frequencies above 10 kmc. The second is ionospheric absorption, arising from collisions between free electrons and molecules. It can be effective below 50 mc.

31.2.2.1. Molecular Absorption. Most of the absorption arising from molecular resonances of the gaseous constituents of the atmosphere is attributed to water vapor and oxygen. There are several absorption bands for each (References 6, 7, and 8) Water vapor has bands centered at 22 kmc, 170 kmc, and some above 300 kmc ($\lambda < 1$ mm). Oxygen's absorption bands in this frequency range center at 60 kmc and 119 kmc.

When attenuation of a signal results from uniform absorption, the signal decreases exponentially with distance. Consequently, the rate of attenuation a is a constant number of decibels per unit distance. Where A = absorption (in decibels), a = decibels per mile, and d = miles,

$$A = ad \tag{31-3}$$

The rate, for two absorptive agents simultaneously present, is additive, and

the absorptive attenuation is to be added to the free-space attenuation (in decibels). When the absorbing agent is not uniformly distributed over the transmission path, the total absorptive attenuation must be integrated over the path.

Naturally the magnitude of the absorption coefficient α depends on the density of the gas through which the wave passes. In the case of oxygen, the density depends almost entirely on altitude only, and so is predictable. The density—and thus also the absorption coefficient α—decreases exponentially with altitude, with a scale height H of about 5 miles; that is, the density has decreased to $1/e$ of its surface value at a height of 5 miles; alternatively, the total oxygen in the atmosphere is the same as would be contained in an atmosphere 5 miles thick having uniform density. Thus the total attenuation A due to oxygen absorption for a wave traveling vertically upward through the entire atmosphere is five times the number of decibels per mile applicable at the surface α_0, for the frequency in question.

For a wave traveling obliquely through the atmosphere, an exact evaluation is a little more involved. However, a reasonably good approximation for ground-to-space propagation can be obtained by assuming an atmosphere of constant density (and thus constant absorptive attenuation rate) and of a thickness equal to the scale height; the effective length of trajectory d through this atmosphere is then calculable geometrically (Figure 31-1).

FIGURE 31-1   Geometry for standard ray calculations

When the take-off angle β is zero (tangent ray), the effective length of path d is given approximately by

$$d = \sqrt{2H} \qquad (31\text{-}4)$$

where d is in miles and H is in feet.

The distribution of water vapor in the atmosphere is not the same as that of oxygen. Water vapor decreases more rapidly with altitude, so that if the concept of scale height were applied it would have a smaller value,—perhaps 2 miles. However, the principal difference lies in the extreme variability of the water content of the atmosphere, which may easily vary by a factor of 3 or 4 at one locality. This means that the decibel value of the absorption can vary by a similar factor. Consequently the figures quoted below must be considered flexible.

The four principal absorption peaks in the frequency range below 300 kmc are listed in Table 31-I. Two of these are for water vapor and two for

TABLE 31-I. GASEOUS MOLECULAR ABSORPTION BANDS

Frequency (kmc)	22	60	119	170
Absorption coefficient (do/mi)	0.20	20	2.7	48
Absorbing agent	H_2O	O_2	O_2	H_2O

oxygen. The absorption coefficients listed apply to conditions at sea level and, in the case of water vapor, are intended to be representative of average temperate climates.

If we take these figures and apply them to four cases of one-way propagation we can obtain an idea of the strength of these absorption bands. The four cases selected for illustrative purposes are (1) vertical propagation from ground to 10 miles height, (2) the same to 300 miles height, (3) oblique propagation, tangent to the ground, out to a height of 10 miles (300 miles slant range), and (4) the same to a height of 300 miles (1800 miles slant range).

TABLE 31-II. ATTENUATION AT ABSORPTION-LINE FREQUENCIES

Frequency (kmc)	22	60	119	170	Transmission path, from earth's surface
Absorption (db)	less than 1	100	14	96	Vertically to 10 mi
Free space Attenuation (db)	146	153	159	162	
Absorption (db)	less than 1	100	14	96	Vertically to 300 mi
Free-space Attenuation (db)	176	183	189	192	
Absorption (db)	29	4600	620	6960	Tangent ray to 10 mi height
Free space Attenuation (db)	176	183	189	192	
Absorption (db)	29	4600	620	6960	Tangent ray to 300 mi height
Free-space Attenuation (db)	191	198	204	207	

Table 31-II lists the total absorption in each of the four cases for the four absorption bands For comparison, the free-space attenuations are also listed, for each of these sixteen categories. Some interesting observations can be made. For transmission vertically upward, the water-vapor absorption at 22 kmc is negligible. At all frequencies (below 300 kmc) the vertical-transmission absorption is less than the free-space attenuation, though it can be very large (on the order of 100 db at the stronger absorption peaks). The total absorption for a tangent ray penetrating the atmosphere is hopelessly large (more than several hundred decibels) for all absorption peaks except the lowest one for water vapor; even here it is appreciable (around 30 db).

At frequencies intermediate between these absorption peaks, the situation is a little more complicated. Each absorption line is broadened by virtue of the thermal motion and collisions of the air molecules. Consequently at intermediate frequencies the bands overlap. However, not only does the density decrease with height, but also the collision broadening decreases. As a result the absorption valleys between the peaks become more exaggerated at increased heights. The three valleys between the four peaks discussed above occur roughly at the frequencies and have the approximate sea-level net absorption coefficients given in Table 31-III.

TABLE 31-III. ABSORPTION MINIMA

Frequency (kmc)	33	80	220
Combined sea-level absorption coefficient (db/mi)	0.11	0.15	6

The lower two of these minima are sufficient to permit oblique penetration through the atmosphere. Picking a few typical frequencies, we may estimate the total attenuation for a tangent ray as in Table 31-IV. From this it can be seen that below 100 kmc the total absorption is overwhelming only in the region around the 60 kmc absorption band. Above 100 kmc it also becomes prohibitive. However, below about 45 kmc, and again between 75 and 100 kmc, it may be tolerable for some applications.

There are other absorbing gases in the atmosphere, but generally their absorption coefficients are only a fraction of a decibel per mile and certainly are overshadowed by water vapor and oxygen. The above discussion is intended to give a general picture of the problem for long paths and for some of the worst cases. In air-to-air or air-to-space situations, the strong absorptions applicable near the earth's surface will be diminished.

31.2.2.2. Ionospheric Absorption. [*] A radio wave propagating through the ionosphere excites the free electrons it encounters into motion.

[*] See References 9, 10, 11, 12, 58.

TABLE 31-IV. TOTAL ABSORPTION OF A TANGENT RAY PENETRATING
EARTH'S ATMOSPHERE

Frequency (kmc)	3	10	22	33	40	50	60	70	80	100
Total Absorption (db)	2.5	3.5	29	20	29	145	4600	131	51	58
Principal Absorbing Agent and Position in Absorption Spectrum			H_2O Max	Min			O_2 Max		Min	

When these electrons collide with air molecules they give up some of their energy to the neutral molecules. In this way, power is abstracted from the passing wave. The attenuation suffered, when the medium is uniform, varies exponentially with distance. If the medium—and thus its absorption coefficient—varies along the transmission path, the total absorption will involve an integration along the path. In the case of ionospheric absorption, the absorption coefficient is proportional to the numbers of free electrons N per unit volume and to the average frequency of collisions ν of an electron with neighboring molecules; it is inversely proportional to the square of the radio frequency f (at frequencies above 10 mc):

$$\alpha \propto N\nu/f^2 \qquad (31\text{-}5)$$

The collisional frequency ν per electron decreases exponentially with height. The electron density N is low near the surface, begins to increase in the D-layer around 40 miles altitude, and takes a marked increase in the E-layer beginning around 60 miles. The product of the two ($N\nu$) reaches a maximum in the 50-mile high region.

Typical magnitudes are such that the absorption is greatest at frequencies which are reflected from the standard ionospheric layers (a few megacycles or so). At a frequency of 30 mc, the absorption in the worst case is rarely more than one decibel. At higher frequencies, because of the inverse frequency-squared relation, ionospheric absorption is entirely negligible.

31.2.3 Line-of-Sight Propagation—Dispersion

A signal propagated through a dispersive medium may suffer distortions owing to the fact that the velocity of propagation is a function of frequency. Thus the phase relationships between high-frequency components and low-frequency components may slip a little, resulting in a distorted waveform. This effect turns out to be a rather academic one, since the dispersion is never sufficiently strong except in cases where other harmful effects such as absorption are predominant. There are only two types of frequency regions where dispersion can be of even academic significance. One is in the neighborhood of the absorption bands above 10 kmc, and the other is in the HF and VHF region affected by the ionosphere.

31.2.3.1. Tropospheric Dispersion. Near each of the molecular resonances, there is an inherent relation between absorption and the rate of change of refractive index with frequency (References 1 and 3) If one considers a narrow pulse (say a fraction of a microsecond), and computes the distance it would have to travel before its higher frequency components have slipped appreciably in phase from its lower frequency components, one finds that the absorption suffered in traveling this distance is far more significant than any pulse distortion suffered. The effects of lateral dispersion in the troposphere are negligible.

31.2.3.2. Ionospheric Dispersion. The refractive index of the ionosphere is inherently a function of radio frequencies (References 11, 12, 13, and 14). Consequently, dispersion is always present to some extent and is not closely linked to absorption. The distortion a pulse or other wideband signal may suffer depends on the integrated electron density along the transmission path and on the frequency. Precise computations are complicated by the fact that the higher frequency components may follow a path slightly different from the lower frequency components. However, reasonably good estimates indicate the following values of minimum pulsewidth which could be transmitted through the entire ionosphere obliquely under strongly ionized conditions without suffering more than a moderate amount of distortion (Table 31-V). For weakly ionized conditions and for transmission

TABLE 31-V. MINIMUM PULSEWIDTH UNDISTORTED BY IONOSPHERIC DISPERSION

Frequency (mc)	50	100	500	1000	5000
Pulse width (μsec)	20	6	0.6	0.2	0.02

vertically upward, the dispersion is less. In this case, the minimum undistorted pulsewidths should be less by a factor of 10 than those shown in Table 31-V.

It is evide... st these pulsewidths are relatively narrow for the associated frequencies. Consequently, dispersion is usually not a serious problem, but it could be for wideband transmissions.

31.2.4 Line-of-Sight Propagation—Refraction

The refractive index of the atmosphere varies with position. When the variation is small in a distance of one wavelength,—as is the case under a good many circumstances,—one may visualize the propagation in terms of rays (References 1 and 3). A spatially varied refractive index gives rise to a bending of the rays. This bending can give rise to an apparent direction of arrival which differs from the true direction in which the source lies. Generally, such deflections in angle are small. It is also possible for rays to follow two or more different trajectories between transmitter and receiver. In this case there may be constructive or destructive interference, resulting in strengthened or weakened signals respectively. Thirdly, there is likely to be a difference in transmission time between the two or more trajectories, when they exist. This multipath delay effectively reduces the bandwidth which the propagation can support and thus leads to distortion of wideband signals (such as narrow pulses).

31.2.4.1. Tropospheric Refraction.[*]   In the lower part of the earth's atmosphere, the refractive index n is determined largely by air pressure P (millibars), temperature T (°K), and humidity (vapor pressure e, millibars) in accordance with Eq (31-6).

$$n = 1 + \left[\frac{79}{T}\left(P + \frac{4800e}{T} \right) \right] \cdot 10^{-6} \qquad (31\text{-}6)$$

Since the refractive index differs from unity by only a few parts in 10^4, it is often convenient to define a quantity N equal to the square brackets in (31-6). Thus

$$N = (n-1) \cdot 10^6 \qquad (31\text{-}7)$$

Typically the surface values of N lie around 300 at sea level. Note that they increase with humidity (e) and decrease with temperature (T). Also they normally decrease with altitude as a consequence of decreasing air pressure p.

The absolute value of tropospheric refractive index is not so important as its gradient, since it is the latter which leads to a bending of the rays. For

[*]See References 1 and 3

the most part, the atmosphere has a marked horizontal stratification, so that the strongest gradients are usually vertical. In these cases, it is the rate of change of n with height that is important. For rays propagating at low angles through a horizontally stratified atmosphere, the radius of curvature r is related to the vertical gradient of refractive index dn/dh by Eq. (31-8):

$$1/r = -(dn/dh) \qquad (31\text{-}8)$$

or

$$10^6/r = -(dN/dh) \qquad (31\text{-}8a)$$

Here the units for r and h are the same. Since the velocity of propagation v is inversely proportional to refractive index

$$v = c/n, \qquad (31\text{-}9)$$

c being the free space velocity of propagation, it follows that a decrease of refractive index with height leads to a downward bending of the ray;—i.e., in Eq (31-8) dn/dh is usually negative ($-dn/dh$ positive) and a positive radius of curvature by definition implies concavity downward.

Under more or less standard conditions n decreases at the rate of 1.2×10^{-6} per hundred feet. The radius of curvature of the earth R leads to a value

$$10^6/R = 4.8 \text{ per } 100 \text{ ft} \qquad (31\text{-}10)$$

The difference in curvature (times 10^6) between a standard ray and the earth's surface is thus

$$(10^6/R) - (10^6/r) = 3.6 \text{ per } 100 \text{ ft} \qquad (31\text{-}11)$$

If we were, artificially, to modify the earth's radius to a value R',

$$R' = (4/3)R$$

then we find that the modified ray curvature $1/r'$, required in order that the difference in curvature remains the same, is

$$1/r' = 0$$

This means that by treating the earth as having $4/3$ times its actual radius, we may consider standard refracted rays as straight lines.

If temperature decreases with altitude less rapidly than in a standard

atmosphere, or if the humidity decreases more rapidly, then the refractive index gradient will exceed the standard value mentioned above, i.e.,

$$- (dN/dh) > 1.2 \text{ per } 100 \text{ ft} \qquad (31\text{-}12)$$

In fact if this negative gradient (rate of *c* ·. *se* of N with height) reaches a value of

$$- (dN/dh) = 4.8 \text{ per } 100 \text{ ft}, \qquad (31\text{-}13)$$

the curvature of a nearly horizontal ray equals the earth curvature. For steeper gradients,

$$- (dN/dh) > 4.8 \text{ per } 100 \text{ ft} \qquad (31\text{-}14)$$

the ray curvature exceeds that of the earth; a ray starting upward at a small angle may be bent around and down back to the surface again. This phenomenon is known as superrefraction and, under other appropriate conditions, as ducting or trapping. It is important for propagation to distances beyond the horizon, as discussed later.

Within line-of-sight, superrefraction is important in several respects. It is usually encountered in connection with horizontally stratified layers of limited vertical extent. Thus there may be a small range of heights within which ray curvatures may exceed that of the earth. This circumstance, together with the detailed nature of the bending makes it possible for rays to be concentrated in a limited height range and even to cross over each other. The greater concentration tends to lead to higher signal levels in the layer at the expense of some regions outside the layer. The crossing of rays leads to an interference which may result in signal enhancement or diminution, depending on the phase. The strong downward bending leads to angles of arrival higher than would be the case under standard conditions. Where rays cross, multipath conditions exist.

The important consideration is the extent and magnitude of these effects. First consider the frequency range over which refractive bending may be a significant influence. At frequencies well below the resonant absorption bands discussed earlier the refractive index is relatively insensitive to frequency. Similar phenomena can occur in the dispersive regions but are more complicated as regards detailed behavior and in any event are likely to be less predominant than considerations related to absorption,—whether a signal can get through to be received in the first place. The lower frequency limit for refractive effects is approached when the radio wavelength begins

to become comparable with the size of the regions over which the refractive
index varies abnormally (for example, with layer thickness). This occurs
below 100 mc. Consequently the frequency range in which refractive effects
are likely to be predominant includes UHF and SHF.

Secondly most of the strong refractive-index gradients are vertical, arising
from horizontal stratification of the air. A wave propagating vertically, nor-
mal to the stratification, is unaffected. As a consequence, the important
refractive effects occur for horizontal or nearly horizontal propagation,—
say for elevation angles of less than 10 degrees.

The most common occurrence of strong refracting effects is an increased
downward bending of the rays. Substandard conditions do occur, in which
rays have a lesser downward curvature than under standard conditions, or
even an upward curvature. However, these conditions are less frequent and
are likely to be associated with air mass instability so that they do not per-
sist or are masked by turbulence. The downward bending concentrates the
radiated power more strongly near the surface,—in and below the layer--
than would otherwise be the case. Consequently, signal strengths are in-
creased on the average. In the theoretical limit, signal power would fall off
inversely as the inverse first power of the distance, rather than the inverse
square characteristic of free-space propagation. This signal enhancement is
rarely achieved in practice. Whatever increase in signal does occur (10 to
15 db) is not so important in line-of-sight situations, in which signal levels
are often adequate anyway, as in beyond-line-of-sight geometries, as dis-
cussed later. However, deep fades associated with multipath interference
are significant. They can result in the temporary loss of signal under con-
ditions in which one might not expect it.

Another phenomenon affecting line-of-sight signal levels is that of radio
holes (Reference 15). This results from the redistribution of power as-
sociated with strongly refracting layers. Unlike absorption, refraction does
not remove power from the transmitted wave. It merely alters the wave's
direction and concentrates the power in some regions at the expense of
others. The regions from which power is diverted have been named radio
holes. They occur most frequently in air-to-air propagation in elevated
layers when both transmitting and receiving terminals are within or near
the layer. Under such conditions it is possible for an otherwise strong signal
to become so weak as to be undetectable. Layers of this sort generally do not
exist at elevations above 10,000 or 15,000 feet. Consequently the radio-hole
phenomenon is unimportant for high-flying aircraft or satellites.

As mentioned above, refractive variations normally give rise to vertical
bending of the rays (References 16, 17, 18, 19, and 20). As a result the
angle of arrival that is more strongly affected is the elevation angle, rather

than the azimuth. Variations in elevation angle of arrival can reach 1 or 2 degrees of departure from the angle for a standard atmosphere. These strongly refracting conditions result from excessive downward bending, so that the ray arrives from above,—i.e., the variations in elevation angle are upward from standard. Rays arriving at angles below that of the standard ray sometimes occur as a result of refraction, but the variations amount to only a few tenths of a degree. Similarly, horizontal variations in angle of arrival are likely to be small, not more than 0.1 or 0.2 degrees. Exceptions may occur in transitory situations such as the passage of a sharp weather front, but these are complicated and unpredictable. It is to be noted that even the strongest of these angular variations are small. In countermeasures applications, they are relatively unimportant except for precise locating techniques.

A consideration which is of more direct concern, however, is that of multipath limitations on bandwidth. When two or more ray trajectories between transmitter and receiver exist, the reciprocal of the difference in transmission time gives a rough approximation to the bandwidth that can be transmitted.* For average refracting layers this bandwidth may be on the order of 40 to 50 mc or so,—adequate for most transmissions. Under more severe conditions the bandwidth may be a factor of 10 smaller than this (i.e., 4 to 5 mc), so that short pulses would be distorted in the transmission.

Tropospheric refraction within line-of-sight is of importance in countermeasures, though not as markedly so as in the transhorizon case. Nevertheless signal levels are affected, so that a jammer situated in a radio hole relative to its intended victim could be thwarted, and a search receiver could be faced with difficulties. Bearing alterations are too minor to hinder direction finding systems or spoofing techniques. However, multipath limitations on bandwidth could interfere with precise signal analysis in some instances.

31.2.4.2. Ionospheric Refraction.† In the ionized regions of the atmosphere, the refractive index depends on the density of free electrons N (per cm^3). For frequencies above 10 mc the refractive index is given to a good approximation by

$$n = \sqrt{1 - \frac{8 \times 10^{-6} \times N(cm^{-3})}{f^2(mc)}} \qquad (31\text{-}15)$$

Note that its values are less than unity and approach zero for sufficiently high electron density N, at any one frequency. Owing to the dispersive

*See Reference 4.
†See References 11, 12, 14, and 21.

nature of the medium, the group velocity of propagation v_g is directly proportional to the refractive index and inversely proportional to the phase velocity v_{ph}:

$$v_g = nc$$
$$= c^2/v_{ph} \tag{31-16}$$

where c is the free-space velocity of propagation. The phase velocity is still inversely proportional to refractive index so that a decrease of n with height leads to a downward bending of a ray.

The downward bending by ionospheric layers is most commonly utilized in HF propagation to distances far beyond the horizon, as discussed later. The important point here is that waves of sufficiently low frequency can suffer total reflection by the ionosphere, even when incident at steep angles. For example, a wave incident vertically on the ionosphere (either from below or from above) will be reflected if the electron density N reaches a value, for any given frequency f, such that the refractive index n drops to zero. For oblique incidence, the electron density need not reach such high values in order that the same radio frequency be reflected; alternatively, a higher frequency ray can be reflected at oblique incidence than at vertical. The law of

$$n \cdot r \cdot \cos \beta = \text{const.} \tag{31-17}$$

refraction is for any one ray, where r is the radial distance from the earth's center and β is the angle with respect to horizontal. For example, as a ray enters the ionosphere from above, the electron density N increases so that the refractive index n decreases according to (31-15); r is decreasing; $\cos \beta$ then must increase, according to (31-17); therefore β decreases. This continues as long as the electron density keeps increasing, so that the ray is bent around until it becomes horizontal, and so is reflected. A similar argument applies to incidence from below.

In practice, frequencies up to 30 mc or slightly higher can at times be reflected in this conventional manner. At 50 mc, the standard E- and F-layers can nearly always be penetrated, but not without causing some deflection to the passing wave. In the intermediate region between 10 mc and 40 mc, where sufficiently steep rays penetrate and others do not, many interesting effects take place. For example, the coverage seen from a point on the earth's surface is somewhat as illustrated in Figure 31-2. A low-angle ray (No. 1) is reflected and returns to earth at some large distance. At a higher angle, some other ray (marked No. 2 in the figure) returns to earth at a minimum distance, defining the skip zone. Ray No. 3 returns to earth
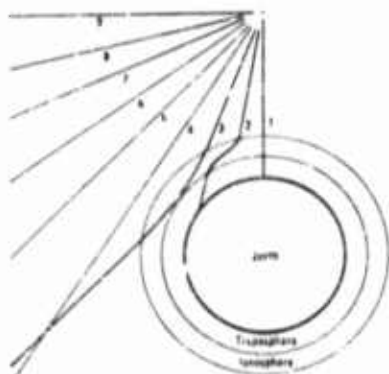
FIGURE 31-2 Ray paths from point on   FIGURE 31-3 Ray paths from point out-
Earth's surface   side atmosphere
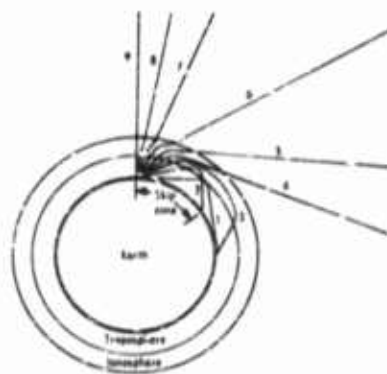
at the same distance as No. 1, causing multipath problems. Ray No. 4 strikes the ionosphere at a sufficiently steep angle to penetrate, as do all higher angle rays at the same frequency. Rays No. 4, No. 5, and No. 6 are deflected in such a manner as to spread their power out very thinly. Rays No. 7, No. 8 and No. 9 are deviated less, so that the field strengths in this cone are approximately the free-space values. This cone of illumination of escaping rays becomes narrower as the frequency is lowered, until it vanishes altogether when the critical frequency of the ionosphere is reached. As one goes up in frequency, of course, this odd type of lens effect introduced by the ionosphere becomes small; in fact it is negligible above about 40 mc.

The coverage as seen from above the ionosphere is illustrated in Figure 31-3, for this same frequency region (10 to 40 mc). Here Ray No. 1, incident vertically, penetrates to the earth's surface. Ray No. 2 is deflected but still penetrates. Ray No. 3 does not penetrate but is reflected at such an angle that it intersects another ray (No. 4) which did not strike the ionosphere. Other rays are unaffected. The cusps occurring at points of intersection of rays such as No. 3 and No. 4 give rise to multipath problems. At lower radio frequencies, a narrower cone of rays about the vertical would penetrate, and more rays would be reflected like No. 3, but at sharper angles. Consequently there would be a larger region with multipath problems.

Above 40 or 50 mc, rays always penetrate but nevertheless may suffer some deflection in angle. The resulting error angle (difference between observed angle of arrival and direct line-of-sight) is greatest when one terminus is imbedded in the ionosphere. Table 31-VI shows some calculated error angles at the ground as a function both of angle above the horizon and of frequency, when the source is in the F-layer at a height of 375 km (234

TABLE 31-VI. ERROR ANGLES (IN DEGREES) AT GROUND FOR SOURCE IN IONOSPHERE

Elevation angle (degrees)	Frequency (mc)			
	50	100	500	1000
1	0.7	0.2	0.03	0.03
7	0.7	0.2	0.02	0.01
30	0.3	0.07	0.005	0.014
60	0.1	0.02	0.001	0.006

miles). It is seen that these errors are not large at 50 mc and above. In fact they are smaller than the resolving powers of most antenna systems at the frequencies involved. At the upper frequency ranges they are smaller than the deviations caused by even mild tropospheric refraction which may easily be an appreciable fraction of a degree.

In its implications for countermeasures, ionospheric refraction is important in several respects. The shadow zone of the earth is distinctly modified. Deflections in bearing, though not large at higher frequencies, are present and can become very large as frequencies descend into the HF region. Multipath can be severe, particularly at lower frequencies. Below the critical frequency, the ionosphere shields the earth from outer space.

31.2.5 Polarization

None of the effects caused by gases in the troposphere influence the polarization of a radio wave. Consequently, when these are the only agents at work, the polarization of a transmitted wave is preserved, and measurement at a receiving point is indicative of the source. Reflections from clouds or the earth's surface can, however, affect the polarization and make reliable determination difficult, since there is no easy way to compute corrections.

The ionosphere exerts a unique influence on the polarization of a wave (References 12, 13, and 22). Because of the presence of the earth's magnetic field, the ionosphere is a bi-refracting medium. These are two values of refractive index and two velocities of propagation. Associated with each is a characteristic polarization. An incident wave generally excites both of these modes. Each travels with a different velocity, so that they emerge from the ionosphere with a phase relationship different from that at their entrance. Thus in recombining they form a different polarization. If initially the polarization was linear, the result of this process is a rotation of the plane of polarization. The amount of rotation depends on the frequency, the angle made with the earth's magnetic field lines and the integrated electron density along the path. As an example of these dependencies, a frequency of 100 mc penetrating through most of the

Ionosphere at an angle nearly parallel with the earth's magnetic field would suffer about a 90-degree rotation under weakly ionized conditions and about 20 complete rotations under strongly ionized conditions. The rotation varies inversely as the square of the frequency, if the path followed is the same.

Two practical consequences follow from this polarization effect. One is that information as to initial polarization is lost. The second is that the arriving polarization may be orthogonal to that of the receiving antenna, (if the latter is linearly polarized) so that the signal is lost,—unless provision is made for an adjustment or an alternate.

31.3 Transhorizon Propagation

Refractive bending, which complicates propagation within the horizon, can under some circumstances be sufficiently strong to carry an appreciable amount of power around the curvature of the earth. In addition, there are other mechanisms which help circumvent the limitations of the horizon. Some of these, such as diffraction and scattering, play a role at both ends of the frequency range being considered. Others, such as reflections from meteor trails and aurora are more specialized.

31.3.1 Transhorizon Propagation—Refraction

The basic refraction phenomena are the same as those discussed for line-of-sight conditions. In the troposphere, they are caused by gradients of temperature and humidity; in the ionosphere, they are caused by layers of free electrons.

31.3.1.1. Tropospheric Refraction.   When the temperature increases with altitude or the humidity decreases with altitude sufficiently rapidly so that the negative refractive index gradient exceeds 4.8×10^{-8} per hundred feet, a ray curvature may exceed that of the earth [see Eq (31-13)]. It is then possible for the ray to be bent downward, reflected from the earth's surface and bent downward again, thus proceeding in a series of hops near the surface. The ray is trapped and follows around the curve of the earth. This phenomenon is known as trapping or ducting, since the waves are effectively guided by the layer (References 1, 3, 23, and 24). This trajectory applies to rays which start off at sufficiently low angles. Rays starting at steeper angles may escape. Figure 31-4 illustrates the ray pattern in a surface duct.

The ray picture however does not tell the whole story, since ray theory gives only an approximation to the true behavior of electromagnetic waves. A more exact solution shows the trapping phenomena to be
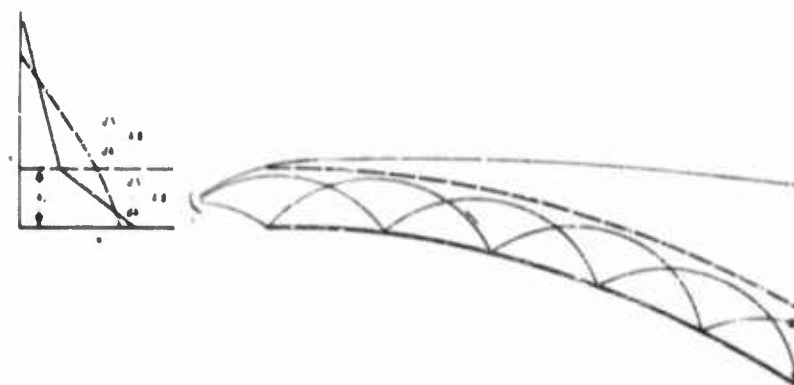
FIGURE 31-4   Duct Propagation

frequency sensitive. Whether an appreciable portion of power is trapped
in the duct depends on the gradient of refractive index, the height inter-
val over which a substantial gradient exists and the radio frequency. To
a certain extent, a duct behaves like a leaky waveguide. Higher fre-
quencies may be effectively trapped while lower frequencies may leak out the
top of the duct for any one layer thickness. While it is difficult to apply
a mathematical formula precisely to this behavior, an indication of the
frequency dependence of trapping can be obtained from Table 31-VII. It

TABLE 31-VII. MINIMUM TRAPPED FREQUENCY AS A FUNCTION OF HEIGHT OF A
SURFACE DUCT

h_o (ft)	20	50	400	1000
f_{min} (kmc)	10	3	0.3	0.05

shows the lowest frequency which can be trapped by a surface duct of
specified thickness. These sample figures are not to be taken too literally,
since the cut-off in frequency is not sharp, and other factors enter such
as steepness of refractive-index gradient, roughness of ground, height
of transmitter and horizontal uniformity of the layer. Nevertheless they
do serve as a guide, and indicate an inverse 3/2-power dependence of
cut-off frequency on duct thickness.

Signal levels in a duct, for frequencies successfully trapped, average
around the free-space level appropriate to the distance covered—even though
the distance may take the transmission well around the curvature of the
earth. Consequently, propagation ranges are increased substantially, some-
times phenomenally, by ducting. Signals have been measured 10 to 15

db above the free-space level at times; at other times deep fades 30 to 40
db below this level are observed. At the trapped frequencies, there is
little gain in signal with increased height of receiver (or transmitter) as
long as it remains within the duct. Just above the duct, signals may be
abnormally low; however, if the duct is not strong or the frequencies are
low enough to be incompletely trapped, the signal level may taper off only
very gradually above the duct.

Situations favoring duct propagation are of various sorts. The most
reliable and extensive trapping conditions occur in the low-latitude high-
pressure areas over the oceans. Here subsidence of the air mass leads
to temperature inversions, and evaporation from the sea surface results
in strong vertical humidity gradients, both favorable to the formation of
ducts. Similar conditions can occur over land but are less consistent,
owing to nonuniformity of most land surfaces and to greater diurnal
variability. A typical, though transitory, type of duct occurs over land
as a result of radiational cooling of the ground on calm nights. In general,
ducts can occur in many areas of the world, but only over low-latitude
ocean areas are they consistent enough to be reliable.

Ducting represents one of the major propagational influences affecting
countermeasures procedures. The increased signal levels permit both jam-
ming and passive detection at greater ranges in surface-to-surface situa-
tions. Variations in angle of arrival are slight enough so as not to disrupt
direction-finding or spoofing techniques seriously,—except possibly in the
vertical plane at low angles. Bandwidth capabilities are likely to be high
enough for reasonably good signal analysis.

31.3.1.2. Ionospheric Refraction.[*]   As mentioned in the discussion
of line-of-sight propagation (Section 31.2.4), ionospheric layers can cause
total reflection of an incident wave, if the electron density is high enough
and the radio frequency low enough. There is a critical frequency, lying
between a few megacycles and about ten megacycles, depending on the
maximum electron density in the ionosphere, below which waves cannot
penetrate[†] and above which they can, if incident at a steep enough angle.
Waves slightly above the critical frequency may be reflected if incident
obliquely, and it is for these frequencies that the phenomenon of skip dis-
tance occurs. (See Figure 31-2 and accompanying discussion). A higher fre-
quency would have a greater skip distance. Consequently, for any one
distance there is a maximum usable frequency (muf), which is the frequency

[*]See References 11, 12, 14, and 21.

[†]An exception occurs in the case of VLF and ELF radiation in the "whistler" mode
(Reference 25).

·   which the distance in question is the skip distance. The quantitative aspects of this reflection process have been well understood for some time; and such organizations as the U. S. National Bureau of Standards issue regular bulletins to facilitate their use in engineering practice. Because of this availability of data, and since we are here concerned primarily with higher frequency ranges, this discussion will be limited in extent and confined to a mention of some aspects of importance to countermeasures.

The largest practical distance on the earth's surface which a wave can reach via one ionospheric reflection is that limited by tangent rays and reflection height. For F-layer reflection at 200 miles height, this distance is around 2500 miles. Greater ground distances can be achieved by multiple hops: two or more ionospheric reflections with intermediate ground reflections.

For frequencies above the critical frequency and at a distance greater than the skip distance, each ionospheric layer can transmit two rays; a high-angle and a low-angle ray (for example No. 1 and No. 3 in Figure 31-2). Each of the two rays may in turn be split into two magnetoionic components by the earth's magnetic field. There may exist as many as two or three reflecting ionospheric layers (Reference 26), the E, F_1 and F_2, at any one time. In addition it is often possible for single—and multi-hop transmission to be present simultaneously. Some of these numerous component rays may be much weaker than others, and therefore not too significant. Nevertheless the detailed picture can be quite complicated. The net combination is likely to result in signal levels not too far from the free-space value. However, the variety of ways these components may combine in phase leads to fading of the signal, which may also be affected by fluctuations in layer height and curvatures. The bandwidth that may be transmitted via any one component is large, but the large differences in transmission time associated with these various components results in a very limited over-all transmission bandwidth. Angles of arrival in the vertical plane may range from nearly zero to 25 or 30 degrees. Horizontally, tilts in the conventional layers may cause departures from the great-circle bearing of a few degrees; — the greatest bearing errors arising from ionospheric propagation are not associated with layer propagation but with irregular types of ionization: aurora, meteors, sporadic-E clouds.

Of these three irregular types, the first two are discussed in separate sections below. However, the third is a layer-type phenomenon and will be considered here (Reference 27). At E-layer heights patches of ionization often exist having high electron densities, small vertical extent and 50 to 100 miles horizontal extent. They are at times capable of reflecting

signals at frequencies above 100 megacycles. Their occurrence is unpredictable and erratic, as the name implies, but when present they can support VHF propagation to distances of 1400 miles. The bandwidth is as wide as a few megacycles unless there are contributions from more than one sporadic-E patch or from other reflecting sources, in which case serious multipath limitations occur.

The polarization of a wave passing through an ionized region depends on the integrated electron density through which the wave passes, regardless of whether the wave is penetrating the region or being reflected from it. Consequently the discussion in Section 31.2.5 is applicable here. For frequencies low enough to be reflected by an ionized layer, all information regarding initial polarization of the wave is lost, except in very special circumstances.

Insofar as conventional reflections from the regular ionised layers are utilised primarily for communications, the countermeasures of importance are jamming of the communications and intercept and analysis of signals. Direction-finding procedures are accurate except where tilting of the layers occurs or sporadic-E clouds cause extreme off-path reflections. In the latter case d-f information can be highly misleading; consequently the possibility of this phenomenon occurring is always a serious concern. Bandwidths are severely limited in most instances by multipath, but this limitation is also imposed on the original transmissions. Sporadic-E, despite its erratic nature, can sometimes be very useful for long range reconnaissance.

31.3.2 Transhorizon Propagation—Scatter

In recent years considerable attention has been given to types of transmission made practical by increases in transmitter powers and by certain operational needs. In the microwave region, these types of transmission cover large ground-to-ground or ground-to-air distances even in the absence of ducts. In the VHF region, transmission above the "maximum" usable frequency is achieved. In both instances, the word "scatter" has become associated with the phenomena, partly because of certain characteristics of the received signal and partly because of postulated mechanisms of propagation. The merits of potential theoretical explanations are less important here than a phenomenological description of experimentally measured characteristics.

31.3.2.1. Tropospheric Scatter.[*] Under nonducting conditions, microwave signal levels drop off quite rapidly once the distance exceeds line-

*See References 5, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

of-sight. There is an interval of distances in which the predominant mechan-
ism by which power reaches the receiver is that of diffraction around the
curvature of the earth. At greater distances another mechanism becomes
predominant as is evident in received signal characteristics. The most out-
standing characteristic is the variation of signal level with distance. Typically
the signal decreases far more slowly with distance than in the diffraction
region. Attenuation rates range from 0.12 to 0.20 decibels per mile—being
perhaps greater at higher frequencies. A useful, though not exact, rule of
thumb for estimating signal levels at frequencies in the general region of
several hundred or a few thousand megacycles, is to consider the signals at
100 miles to be 60 decibels below the free-space value and to be decreasing
at 0.17 decibels per mile. This applies to that portion of the path between
horizons. From each antenna to its horizon, the transmission loss is com-
puted at the free-space value. In this way the effect of elevated and
airborne antennas can be taken into account. More detailed computational
procedures are given elsewhere.

The net result is that with high powered transmitters, sensitive receivers
and substantial antenna gains, large distances can be achieved. One to four
hundred miles are common. Ground-to-ground signals have been received at
400 mc out to 800 miles, and ground-to-air signals at 200 mc to 1000
miles, though the latter, at least, appear to be attributed to ionospheric
scatter.

For a fixed path the signal is characterized by fairly rapid and extensive
fading. The fading range between levels exceeded 10 percent of the time
and 90 percent of the time is 15 to 20 decibels. Fading rates are frequency-
dependent. Often, but not always, they are directly proportional to
frequency and run around 1 cps at 3000 mc. Over short time periods
the fading is often found to have a Rayleigh distribution.

There is ample evidence to indicate that the signal does not always
arrive exactly along the great-circle route. It is largely this characteristic
which suggests a scattering mechanism as the explanation for this type
of propagation. On the average, this scattering of the radio wave spreads
the power over a range of angles one or two degrees in extent. There are
three immediate consequences, all interrelated. One is an apparent broad-
ening of a narrow beam transmission, or a distorting of an antenna beam
which is narrower than the scattering pattern of the atmosphere. The
second is a degradation of transmission bandwidth; the multipath associated
with off-path scattered components introduces delays which effectively
reduce the usable bandwidths to a few megacycles. The third is an apparent
loss in antenna gain. A large antenna having a sufficiently narrow beam may
not receive power scattered from off-path angles; consequently, the signal

It receives relative to a smaller broader-beamed antenna is not so great as the difference in their plane-wave gains would indicate. The magnitude of this "gain-loss" has been measured to be a few decibels for beamwidths of less than one degree. Its variation with distance, frequency and meteorological conditions has not always presented a consistent picture, however.

In view of the relatively limited number of careful measurements in this new field, many of these statements have to be accepted with some reservations. The measurements indicate the off-path components may at times be a... in number but move around rapidly. Thus the beam broadening is not necessarily observed instantaneously, and instead there is a fluctuation in angle of arrival over a range of a few degrees (Figure 31-5). Other experiments indicate that conditions at a few hundred
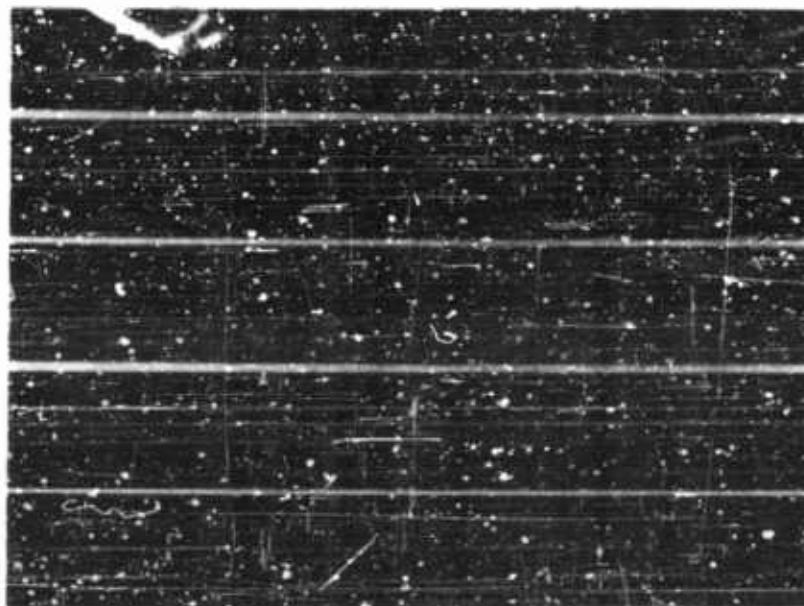
FIGURE 31-5 Fluctuations in transhorizon angle of arrival. Each trace corresponds to a 4-degree sector scan from left to right, the center representing the great-circle bearing to the distant transmitter. Successive scans (traces) are 0.1 second apart. The signal peaks indicating angle of arrival are seen to vary rapidly, at times splitting into several simultaneous components.

megacycles often differ from those at a few thousand bandwidths are greater, gain degradation less, and fading rates slower than is consistent

with the above picture. Finally there is appreciable difference between the measurements made at one time and place and those at others. Caution should be used in applying formulas derived from any one theoretical model.

Transhorizon tropospheric propagation has applications in reconnaissance. Generally there is too much loss in getting beyond the diffraction region to make power jamming feasible and too much variability to make deceptive jamming realistic. But the gradual attenuation of signal level with distance makes possible the reception of signals at large ranges and at frequencies too high to utilize ionospheric reflections. In this application, there may be distortion of short (less than one microsecond) pulses or other wideband transmission, and of antenna beam shapes. There is also an uncertainty in angle of arrival (one or two degrees) which is unimportant except for accurate direction-finding systems. These are limitations, but not serious ones.

31.3.2.2. Ionospheric Scatter.[*] At frequencies between 30 and 100 mc, signals can be transmitted via the ionosphere over distances ranging from 500 to 1400 miles. The mechanism of propagation here is related to an irregular ionospheric structure at E-layer heights (60 miles) attributable to turbulence or to large numbers of small meteors. It is thus a scattering process and is characterized by a diversified angle of arrival, antenna gain degradation in the case of narrow beams, rapid fading of the signal and limited bandwidth owing to multipath. Although the general nature of the signal characteristics is somewhat similar to the tropospheric case, there are many detailed differences. Since this type of scatter is limited to E-layer heights, there is a cut-off at short distances (less than 500 miles) where the scatter angle becomes too great. The scatter is caused by unevenly distributed free electrons and appears to vary roughly as the inverse seventh power of the frequency, on the average. The usable bandwidth is quite narrow:—several kilocycles or so.

Countermeasures applications are largely in the area of signal detection, as in the case of tropospheric, and for the same reasons: signal levels are low. The limitations are similar, but more severe in the case of transmissable bandwidth. Distances in general are greater, owing to the greater height of the scattering region, but the frequency range is far more limited. By way of contrast, the troposphere-scatter frequency range extends downward into the region below 100 mc where it is still operative at shorter distances but becomes masked by competing mechanisms. On

[*]See Reference 36

the higher frequency side, ionospheric scatter dies off rapidly, and tropospheric scatter becomes predominant; the latter's upper limit in frequency is set by absorption (above 10 kmc) not by failure of the mechanism. There is a region, then, between 30 and 500 mc, (to choose fairly broad limits) in which both tropospheric and ionospheric scatter may be present simultaneously at large distances. Another overlap in scatter mechanisms occurs in the case of meteor trails, which is the subject of the next section.

31.3.3 Propagation via Meteor Trails*

When a meteor enters the atmosphere it usually leaves a trail of ionization somewhere in the height range of 50 to 60 miles. Free electrons in this trail can scatter incident radio waves during the brief interval (fraction of a second to several seconds) before they recombine with the molecules. The scatter for any one meteor trail is directive, depending on the orientation of the trail. Consequently, the details of meteor scatter are involved and can best be studied statistically.

The frequency ranges over which meteor scatter is effective vary from the HF region, where it is overlapped by conventional layer reflections to a few hundred megacycles. At the lower frequency ranges, and for backscatter, the meteor-trail echo is characterized by a rapid rise and slower exponential decay. Size and duration depends on the individual meteor trail. On the average, the peak signal strength in this echo varies inversely as the cube of the frequency, while the duration of the echo varies inversely as the square of the frequency. Thus the average power scattered by meteors varies inversely with the fifth power of the frequency. In going to higher frequencies, and oblique paths, the nature of the meteor echo changes to a more nearly symmetrical rise and fall. The upper frequency limit is set by the strong frequency dependence mentioned above and by the powers and system sensitivities available. Higher powers make use of the fainter meteor trails which otherwise would not be seen. There are large numbers of these small meteors. Indeed one viewpoint in regard to ionospheric scatter is that the required inhomogeneous structure of the atmosphere is caused by the constant overlap of numerous small meteor trails each one of which would be too feeble to be detected individually.

The stronger trails are not continuously present. Consequently propagation via these trails proceeds in irregular, short-lived bursts. Since the trails occur somewhat randomly, with different orientations, the particular trajectory taken by a scattered radio wave varies from one meteor to the next. Consequently, the direction of arrival is not constant, and meteor-trail

*See References 40, 41, 42, 43, 44, 45, 46, 47, 48, and 49.

propagation does not provide an accurate method of direction finding unless more sophisticated techniques are used to obtain supplementary information. The bandwidth that can be transmitted via any one meteor burst is very large (many megacycles). During the later stages of a long trail, as it gets distorted by wind shear, the bandwidth is reduced, owing to multipath from different parts of the trail, but is still appreciable (a megacycle or so). However, the presence of more than one meteor in the beam simultaneously leads to serious multipath limitations (several kilocycles). In addition the motion of the meteor trail causes a doppler shift, which is not very large in magnitude, being as much as 20 cps at 50 mc and proportional to frequency.

Meteor trails have a limited application in countermeasures. They can be useful in signal detection in that moderate powers can be detected at large distances (1500 miles ground-to-ground). However, the transmission is sporadic and directional information poor. Their use in power jamming is negligible. However, one application of meteor transmission has been in a burst-communication system, which is attractive because of the security aspects associated with the directive properties of meteor scatter. The possibility of confusing such a communications system by spoofing techniques is not to be overlooked.

31.3.4 Propagation via Aurora*

High energy particles shot from the sun are deflected by the earth's magnetic field to enter the upper reaches of the atmosphere at high magnetic latitudes. The resulting ionization takes on a variety of forms. However, in all cases the columns of ionization line up along the earth's magnetic-field lines, and reflection takes place when the incident and reflected rays are perpendicular to the field lines. This geometrical limitation, when combined with the fact that auroral forms rarely penetrate below 60 miles altitude, severely restricts the regions from which auroral reflections can be obtained. For example, they are not detected above 65-degrees latitude looking south. On the other hand, at a latitude of 60 degrees looking poleward they can be detected over an arc 120 degrees in width.

The reflection coefficients actually are small, on the order of 10^{-8}. Correspondingly, absorption associated with aurorae is negligible, being one db or so at 30 mc and decreasing rapidly with frequency. As a result aurorae do not appreciably disrupt straight-through propagation. However, they are of sufficient magnitude for easy detection of reflections at fre-

*See References 30, 51, 52, 53, and 54.

quencies around 50 mc. In going upward in frequency, the reflected signals decrease in amplitude. The law of decrease with frequency is not always consistent, varying from an inverse third to an inverse seventh power. However, current equipments permit auroral detection at 400 mc without difficulty.

Reflected signals are characterized by high fading rates and strong doppler shifts more than ten times those observed in meteor scatter. Both vary proportionately with frequency. In regions where auroral detection is possible, the large spread in angle permitted by the above mentioned geometrical restrictions results in a severe multipath problem. Consequently, signal fidelity is extremely unreliable.

It is clear from all these characte. tics that auroral propagation has definitely limited possibilities in countermeasures. Its best use in any event would be in conjunction with a monitor radar to tell when an adequate aurora is present. The principal advantages are its existence at times when other mechanisms may not suffice and the reasonably good signal strengths obtainable.

31.3.5 Diffraction

Because of the wave nature of propagation, the shadow region around the earth's bulge or behind mountains is not completely devoid of signal, even in the absence of refraction, reflection, scattering, or other deflecting causes. Waves penetrate the shadow regions by diffraction. The magnitude and distribution of the diffraction fields depend on the geometry involved and on the wavelength (References 55 and 56).

For diffraction around a spherical earth, the calculations are quite complicated. They have been worked out and presented in convenient usable form (References 5 and 56). At low frequencies, this ground wave is of considerable importance. As one goes up in frequency, the rate at which the diffraction field decreases with distance beyond line-of-sight goes up rapidly. Thus at microwaves, there is a relatively short distance interval in which diffraction fields predominate. It lies in the region just beyond line-of-sight; the diffraction fields decrease so rapidly with distance that they soon drop below the scatter-field level.

The situation is altered if the rounded earth's horizon is replaced by a relatively sharp mountain range (Reference 57). This "knife-edge" diffraction results in much stronger fields in the shadow region. Consequently diffraction outweighs scatter over a greater distance range. This signal enhancement has been given the somewhat paradoxical name of "obstacle-gain." It can be utilized under some specific situations. Signal levels tend to be steady but depend on details of the geometry and the interrelationship

with refracting conditions. Since very few horizons present a true knife-edge profile, the occasions on which one should expect to find obstacle gain are not always clear. The phenomenon has definitely been shown to exist, but more measurements are needed on fading rates and transmission bandwidth.

In countermeasures applications, diffraction fields can provide a very useful supplement to mechanisms for signal detection in special situations. The application in jamming techniques is far more limited. Direction finding accuracy and signal distortion may be a problem.

31.3.6 Artificial Modifications of the Propagation Medium

People are always alert to the possibility of artificially inducing changes in the atmosphere which might extend or reduce transmission ranges, depending on the purpose to be served. It must be borne in mind in this regard that the atmosphere is very large compared with most man-made disturbances, so that it is necessary either to cause an unusually large disturbance or to find a trigger mechanism to which natural changes in the atmosphere are sensitive.

Attempts to find absorbing gases which would serve as radar camouflage have not proved overly practical. Similarly a stimulation of tropospheric scatter by sound waves or blast has not been achieved.

In the case of the ionosphere, the situation is a little better. Release of chemicals at E-layer heights introduces material easily ionized by solar radiation. The result is a small cloud of ionization which persists for a short time and from which reflections can be obtained. This technique has potential applications in reconnaissance but has not been thoroughly explored at this writing.

Nuclear explosions create a large enough disturbance to have considerable effect. The greatest effects are caused by high level explosions and seriously influence the ionosphere. Ionization at 60-mile heights leads to markedly increased absorption at HF. Artificial aurorae are produced from which reflections can be obtained above 500 mc. Undoubtedly much more research will be done in investigating this phenomenon.

31.3.7 Transhorizon Ranges

By way of quick summary, and at the risk of gross oversimplification, Figure 31-6 is included to give a rough indication of ranges that can be achieved at various frequencies for various types of propagation. This figure is applicable to ground-to-ground transmission beyond line-of-sight. This information must be utilized in the light of the qualifications and variability described in the text.
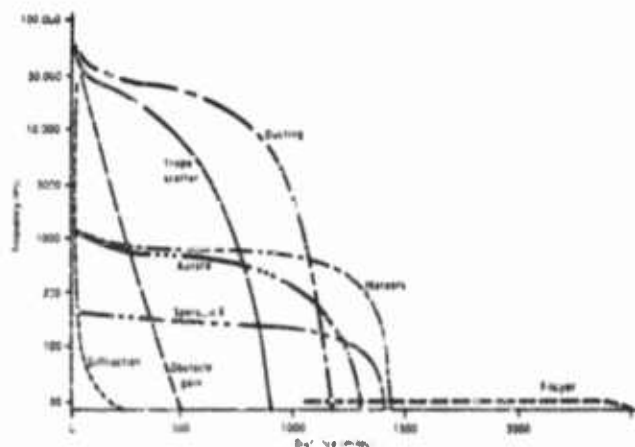
FIGURE 31-6   Maximum ground-to-ground propagation ranges

REFERENCES

1. Burrows, C. R. and S. S. Attwood, "Radio Wave Propagation," Academic Press, Inc., New York, 1949.

2. International Telephone and Telegraph Corporation, "Reference Data for Radio Engineers," ITT, New York, 1956.

3. Kerr, D. E., "Propagation of Short Waves, "McGraw-Hill Book Company, Inc., New York, 1951.

4. Terman, F. E., "Electronic and Radio Engineering," McGraw-Hill Book Company, Inc., Chapt. 22, 1955.

5. Norton, K. A., "Transmission Loss in Radio Propagation: II," National Bureau of Standards Report 5092, Boulder, Colorado.

6. Bean, B. R. and R. Abbott, Oxygen and Water Vapor Absorption of Radio Waves in the Atmosphere, *Review Geofisica Pura E Applicata*, Milano, Vol. 37, pp 127-144, 1957.

7. Rogers, T. E., "Calculated Centimeter-millimeter Water Vapor Absorption at Ground Level," Proceedings of Conference on Radio Meteorology at the University of Texas, November 9-12, 1953.

8. Tolbert, C. W. and A. W. Straiton, Experimental Measurement of the Absorption of Millimeter Radio Waves over External Ranges, *IRE Trans. on Ant. and Prop.*, Vol. AP-5, pp. 239-241, April 1957.

9. Farmer, F. T. and J. A. Ratcliffe, Measurements of the Absorption of Wireless Waves in the Ionosphere, *Proc. Roy. Soc.*, Vol. A 151, pp. 370-383, 1935.

10. Martyn, D. F., The Propagation of Medium Radio Waves in the Ionosphere, *Proc. Phys. Soc.*, Vol. 47, p.p. 323-329, 1945.

11. Mitra, S. K., The Upper Atmosphere, *The Asiatic Society*, Calcutta India, 1952.

12. Ratcliffe, J. A., The Physics of the Ionosphere, *Proc. Phys. Soc.* (London), 1955.

13. Appleton, E. V., Wireless Studies of the Ionosphere, *J. Inst. Elec. Engrs.*, Vol. 71, pp. 642-650, 1932.

14. Pedersen, P. O., The Propagation of Radio Waves, *Danmarks Naturvidenshabelige Samfund*, Copenhagen, 1927.

15. Wong, M. W., Refraction Anomalies in Airborne Propagation, *Proc. IRE*, Vol. 46, pp. 1633-1638, September 1958.

16. Counter, V. A., "Propagation of Radio Waves Through the Troposphere and Ionosphere," Lockheed Aircraft Corporation Report LMSD-2066-R1, May 1958.

17. Crawford. A. B. and W. C. Jakes, "Selective Fading of Microwaves," *Bell System Tech. J.*, Vol. 31, pp. 68-90, January 1952.

18. Crawford, A. B. and W. M. Sharpless, Further Observations of the Angle of Arrival of Microwaves, *Proc. IRE*, Vol. 34, pp. 845-84s, November 1946.

19. Straiton, A. W., Microwave Phase Front Measurements for Overwater Paths of 12 and 33 Miles, *Proc. IRE*, Vol. 37, pp. 808-813, July 1949.

20. Straiton, A. W. and J. R. Gerhardt, Results of Horizontal Microwave Angle-of-arrival Measurements by the Phase Difference Method. *Proc. IRE*, Vol. 36, pp. 916-922, July 1948.

21. Ionospheric Radio Propagation, *Natl. Bur. Standards Circ.*, 1948.

22. Hedlund, D. A. and L. C. Edwards, Polarization Fading Over an Oblique Incidence Path, *IRE Trans. on Ant. and Prop.*, Vol. AP-6, pp. 21-25, January 1958.

23. Gossard, E. E. and L. J. Anderson, The Effect of Superrefractive Layers on 50-5000 Mc. Nonoptical Fields, *IRE Trans. on Ant. and Prop.*, Vol AP-4, pp. 175-178, April 1956.

24. Katzin, M., R. Bauchman, and W. Binnian, 3- to 9- Centimeter Propagation in Low Ocean Ducts, *Proc. IRE*, Vol 35, pp. 891-906, September 1947.

25. Storey, L. R. O., An Investigation of Whistling Atmospherics, *Phil. Trans. Roy. Soc. London*, Vol 246, pp 113-141, July 9, 1953.

26. Ionosphere-IGY Issue, *Proc. IRE*, Vol. 47, February 1959.

27. Smith, E. K., Jr., A Study of Sporadic-E on a World-Wide Basis, National Bureau of Standards Report 1875, December 1955.

28. Booker, H. G and J. T. deBettencourt, Theory of Radio Transmission by Tropospheric Scattering Using Very Narrow Beams, *Proc. IRE*, Vol 43, pp. 281-290, March 1955.

29. Booker, H. G. and W. E. Gordon, A Theory of Radio Scattering in the Troposphere, *Proc. IRE*, Vol. 38, pp. 401-412, April 1950.

30. Chisholm, J. H., et al., Investigations of Angular Scattering and Multipath Properties of Tropospheric Propagation of Short Radio Waves Beyond the Horizon, *Proc. IRE*, Vol. 43, pp. 1317-1335, October 1955.

31. Doherty, L. H. and G. Neal, A 215-Mile 2720 Mc Radio Link, *IRE Trans. on Ant. and Prop.*, Vol. AP-7, pp. 117-126, April 1959.

32. Friis, H. T., A. B. Crawford, and D. C. Hogg, A Reflection Theory for Propagation Beyond the Horizon, *Bell System Tech. J.*, Vol. 36, May 1957.

33. Gordon, W. E., Radio Scattering in the Troposphere *Proc. IRE*, Vol. 43, pp. 23-28, January 1955.

34. Scatter Propagation Issue, *Proc. IRE*, Vol. 43, October 1955.

35. Megaw, E. C. S., Scattering of Electromagnetic Waves by Atmospheric Turbulence, *Nature, London*, Vol. 166, pp. 1100-1104. December 30, 1950.

36. Norton, K. A., Transmission Loss in Radio Propagation, *Proc. IRE*, Vol. 41, pp. 146-152, January 1953.

37. Waterman, A. T., Jr., A Rapid Beam-swinging Experiment in Transhorizon Propagation, *IRE Trans. on Ant. and Prop.*, Vol. AP-6.4, pp. 338-340, October 1958.

38. Waterman, A. T., Jr., Some Generalized Scattering Relationships in Transhorizon Propagation, *Proc. IRE*, Vol. 46, pp. 1842-1848, November 1958.

39. Bailey, D. K., R. Batemen, et al., A New Kind of Radio Propagation at Very High Frequencies Observable Over Long Distances, *Phys. Rev.*, Vol. 86, pp. 141-145, April 15, 1952.

40. Eshleman, V. R., Theory of Radio Reflection from Electron-ion Clouds, *IRE Trans. on Ant. and Prop.*, Vol. AP-3, pp. 32-39, January 1955.

41. Eshleman, V. R. and L. A. Manning, Radio Communication by Scattering from Meteoric Ionization, *Proc. IRE*, Vol. 42, pp. 530-536, March 1954.

42. Eshleman, V. R., P. B. Gallagher, and R. F. Mlodnosky, "Meteor Rate and Radiant Studies" Stanford University Final Report, Contract AF 19(604)-1031, California, February 1957.

43. Forsyth, P. A., E. L. Vogan, D. R. Hansen, and C. O. Hines, The Principles of Janet — a Meteor-burst Communication System, *Proc. IRE*, Vol. 45, pp. 1642-1657, December 1957.

44. Meteor Papers Issue, *Proc. IRE*, Vol. 45, December 1957.

45. McKinley, D. W. R. and A. G. McNamara, Meteoric Echoes Observed Simultaneously by Back-scatter and Forward-scatter, *Can. J. Phys.*, Vol. 34, pp. 625-637, July 1956.

46. Manning, L. A. and V. R. Eshleman, Meteors in the Ionosphere, *Proc. IRE*, Vol. 47, pp. 186-199, February 1959.

47. Villard, O. G., Jr., V. R. Eshleman, L. A. Manning, and A. M. Peterson, The Role of Meteors in Extended-range VHF Propagation, *Proc. IRE*, Vol. 43, pp. 1473-1481, October 1955.

48. Villard, O. G., A. M. Peterson, L. A. Manning, and V. R. Eshleman, Extended-range Radio Transmission by Oblique Reflection from Meteoric Ionization, *J. Geophys. Research*, Vol. 58, pp. 83-93, March 1953.

49. Vincent, W. R., R. T. Wolfram, B. M. Sifford, W. E. Jaye, and A. M. Peterson, Analysis of Oblique Path Meteor-Propagation Data From the Communications Viewpoint, *Proc. IRE*, Vol. 45, pp. 1701-1707, December 1957.

50. Bowles, K. L., Doppler Shifted Radio Echoes from Aurorae, *J. Geophys. Research*, Vol. 59, pp. 553-555, December 1954.

51. Leadabrand, R. L., R. I. Presnell, M. R. Berg, and R. B. Dyce, Doppler Investigations of the Radar Aurora at 400 Mc, *J. Geophys. Research*, Vol. 64, No. 11, September 1959. (to be published)

52. Leadabrand, R. L. and A. M. Peterson, Radio Echoes from Auroral Ionization Detected at relatively Low Geomagnetic Latitudes, *IRE Trans. on Ant. and Prop.*, Vol. AP-6, pp. 65-79, January 1958.

53. Little, C. G., W. M. Rayton, and R. B. Roof, Review of Ionospheric Effects at VHF and UHF, *Proc. IRE*, Vol. 44, pp. 992-1018, August 1956.

54. Nichols, B., Auroral Ionization and Magnetic Disturbances, *Proc. IRE*, Vol. 47, pp. 245-254, February 1959.

55. Bremmer, H., "Terrestrial Radio Waves," Elsevier Publishing Company, Inc., New York, 1949.

56. Norton, K. A., The Calculation of Ground-Wave Field Intensity Over a finitely Conducting Spherical Earth. *Proc. IRE*, Vol. 29, pp. 623-639, December 1941.

57. Kirby, R. S., H. T. Dougherty, and P. L. McQuate, Obstacle Gain Measurements Over Pikes Peak at 60 to 1,046 Mc, *Proc. IRE*, Vol. 43, pp. 1467-1472, October 1955.

58. Brandstatter, J. J., *The Propagation of Waves and Rays in Plasmas*, McGraw-Hill Book Co., New York, 1961.

Index